

Chapter 1

PROC ANOM and General Statements

Chapter Contents

OVERVIEW	5
Uses of Analysis of Means	5
Terminology	6
History	7
Using the ANOM Procedure	7
SYNTAX OVERVIEW FOR THE ANOM PROCEDURE	8
BY and ID Statements	8
Graphical Enhancement Statements	9
SYNTAX FOR THE PROC ANOM STATEMENT	9

PROC ANOM and General Statements

Chapter 1

PROC ANOM and General Statements

Overview

Analysis of means (ANOM) is a graphical and statistical method for simultaneously comparing k treatment means with their overall mean at a specified significance level α . You can use the ANOM procedure to create ANOM charts for various types of response data, including continuous measurements, proportions, and rates.

In addition, you can use the ANOM procedure to

- create charts from either response values or summarized data
- analyze multiple response variables
- specify decision limits in terms of the significance level (α)
- compute decision limits from the data and automatically adjust decision limits for unequal sample sizes
- save chart statistics and decision limits in output data sets
- tabulate chart statistics and decision limits.

Uses of Analysis of Means

Many statistical quality improvement applications involve a comparison of treatment means to determine which are significantly different from the overall average. For example, a manufacturing engineer might run an experiment to investigate which of six positions on a machine are producing different output, in the sense that the average measurement for each position differs from the overall average. Likewise, a health care system administrator might ask which clinics in the system have a higher or lower rate of admissions than the average for all clinics.

Questions of this type can be answered with *analysis of means*, which is an alternative to one-way analysis of variance (ANOVA) for a fixed effects model. However, unlike ANOVA, which simply determines whether there is a statistically significant difference in the treatment means, ANOM identifies the means that are significantly different. As a statistical technique, ANOM is a method for making multiple comparisons that is sometimes referred to as a "multiple comparison with the weighted mean". Analysis of means lends itself to quality improvement applications because it has a simple graphical representation that is similar to a Shewhart chart and requires little training to interpret. This representation is also useful for assessing practical significance.

Figure 1.1 illustrates a typical ANOM chart. The central line represents the overall average. The treatment means, plotted as deviations from the overall average are

PROC ANOM and General Statements

compared with upper and lower decision limits to identify which are significantly different from the overall mean (in this case, the means corresponding to the first, fourth, and sixth positions).

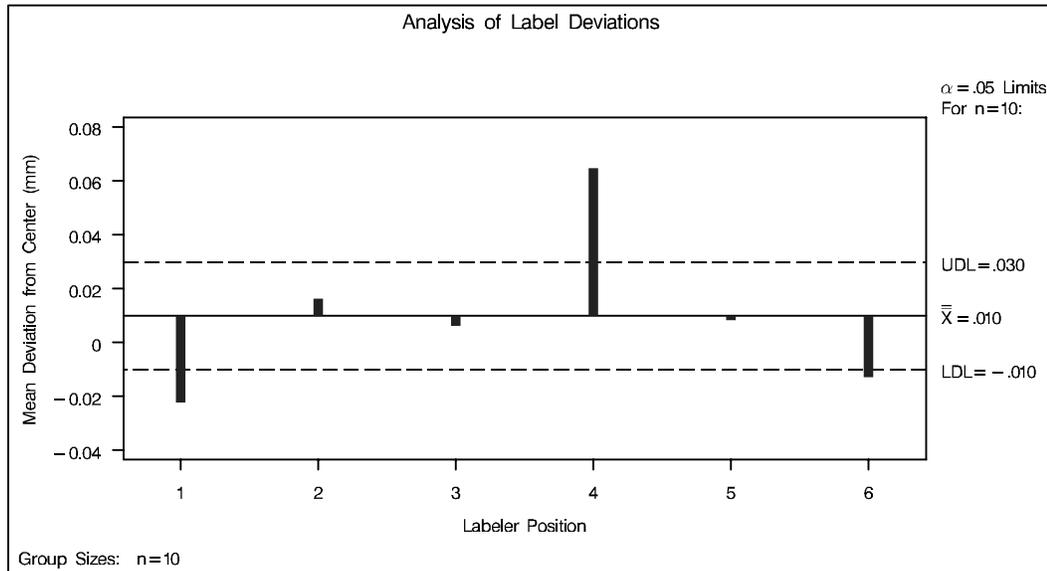


Figure 1.1. Typical ANOM Chart

Although the term "analysis of means" suggests that the method is intended for means of continuous response measurements, the method is also applicable to means of attributes data, including proportions and rates.

Analysis of means was introduced as a tool for statistical quality control by Ellis Ott in 1967, and it became popular during the early 1980s, when it was applied to experimental data in manufacturing. In this setting, measurements are taken at a number of treatment levels (factor levels). During the 1990s, the use of ANOM spread to service industry applications and, in particular, health care quality improvement. In these settings, data (such as utilization rates) are observed for a number of groups (such as hospitals or clinics).

Terminology

In order to accommodate the growing variety of modern applications for analysis of means, the term *group* is used instead of treatment level throughout the documentation for the ANOM procedure. Likewise, the term *group-variable* is used to refer to the variable in the input data set that classifies the observations into treatment levels. In the ANOM procedure, a *group-variable* plays the same role as a CLASS variable in the GLM and ANOVA procedures, and it is syntactically the same as a *subgroup-variable* in the SHEWHART procedure.

The nomenclature for ANOM charts is the same as that for Shewhart charts: \bar{X} charts for means, p charts for proportions, and u charts for rates. Consequently, the syntax for the ANOM procedure is patterned after the syntax for the SHEWHART proce-

ture. However, there are some important differences between ANOM charts and Shewhart charts:

- Analysis of means is formally a test of hypothesis, whereas a Shewhart chart is used to distinguish between special and common causes of variation.
- In an ANOM chart, the horizontal axis corresponds to the *group-variable*, and it identifies the groups, which can be displayed in any order. In a Shewhart chart, the horizontal axis corresponds to the *subgroup-variable*, and it identifies the order in which the subgroup measurements were taken.
- An ANOM chart displays response summary statistics for a set of groups (treatments) at a specific time. A Shewhart chart displays subgroup summary statistics for a specific process where the subgroups are made up of measurements taken over successive points in time.
- In an ANOM chart, the decision limits are determined by a specified significance level (α), which is the probability that under the null hypothesis of no treatment differences, at least one of the response summary statistics will exceed the decision limits. In a Shewhart chart, control limits are typically computed as 3σ limits.

History

Analysis of means compares the absolute deviations of group means from their overall mean, an approach that was initially studied by Laplace in 1827. Halperin and others derived a version of this method in the form of a multiple significance test in 1955. Ott developed a graphical representation for the test and introduced the term “analysis of means” in 1967. Refer to Ott (1967) and Ott (1975).

P. R. Nelson (1982a) and L. S. Nelson (1983) provided exact critical values for ANOM when the groups have equal sample sizes. P. R. Nelson (1991) developed a method for computing exact critical values for ANOM when the group sample sizes are not equal. Refer to Nelson, Coffin, and Copeland (2003) for more information on the use of ANOM in engineering experimentation.

Using the ANOM Procedure

The PROC ANOM statement invokes the ANOM procedure and it optionally identifies various data sets.

To create an ANOM chart, you specify a chart statement (after the PROC ANOM statement) that specifies the type of ANOM chart you want to create and the variables in the input data set that you want to analyze. For example, the following statements request a basic ANOM chart for treatment means:

```
proc anom data=values;
  xchart weight*treatment;
run;
```

PROC ANOM and General Statements

Here, the DATA= option specifies an input data set (*values*) that contains the *response* measurement variable (*weight*) and the *group-variable* (*treatment*). You can use options in the PROC ANOM statement to

- specify input data sets containing variables to be analyzed, decision limits, and annotation information
- specify a graphics catalog for saving graphical output

Note: If you are learning to use the ANOM procedure, you should read both this chapter and the “Getting Started” section in the chapter for the chart statement that corresponds to the chart you want to create.

Syntax Overview for the ANOM Procedure

The following are the primary statements that control the ANOM procedure:

```
PROC ANOM < options > ;
  XCHART (responses)/*group-variable
    < (block-variables) > < =symbol-variable > < / options >;
  PCHART (responses)/*group-variable
    < (block-variables) > < =symbol-variable > < / options >;
  UCHART (responses)/*group-variable
    < (block-variables) > < =symbol-variable > < / options >;
  BOXCHART (responses)/*group-variable
    < (block-variables) > < =symbol-variable > < / options >;
  INSET keyword-list < / options >;
```

The PROC ANOM statement invokes the procedure and specifies the input data set. The chart statements create different types of charts. You can specify one or more of each of the chart statements. For details, read the chapter on the chart statement that corresponds to the type of chart that you want to produce.

BY and ID Statements

In addition, you can optionally specify one of each of the following statements:

BY variables ;

ID variables ;

The BY statement specifies variables in the input data set that are used for BY processing. A separate chart is created for each set of observations defined by the levels of the BY variables. The input data set must be sorted in order of the BY variables.

The ID statement specifies variables used to identify observations. The ID variables must be variables in the DATA= or SUMMARY= input data sets.

The ID variables are used in the following ways:

- If you create an OUTSUMMARY= or OUTTABLE= data set, the ID variables are included. If the input data set is a DATA= data set, only the values of the ID variables from the first observation in each group are passed to the output data set.
- If you specify the TABLEID or TABLEALL options in a chart statement, the table produced is augmented by a column for each of the ID variables. Only the values of the ID variables from the first observation in each group are tabulated.
- If you specify the BOXSTYLE= SCHEMATICID option or the BOXSTYLE= SCHEMATICIDFAR option in the BOXCHART statement, the value of the first variable listed in the ID statement is used to label each extreme observation.

Graphical Enhancement Statements

You can use TITLE, FOOTNOTE, and NOTE statements to enhance graphical and printed output. You can also use AXIS, LEGEND, and SYMBOL statements to enhance your charts. For details, refer to *SAS/GRAPH Software: Reference* and see the chapter for the chart statement that you are using.

Syntax for the PROC ANOM Statement

The syntax for the PROC ANOM statement is as follows:

PROC ANOM *options* ;

The PROC ANOM statement starts the ANOM procedure and optionally identifies various data sets. The following options can appear in the PROC ANOM statement.

ANNOTATE=*SAS-data-set*

ANNO=*SAS-data-set*

specifies an input data set containing ANNOTATE= variables as described in *SAS/GRAPH Software: Reference*. You can use this data set to add features to ANOM charts. Features provided in this data set are displayed on every chart produced in the current run of the ANOM procedure.

BOX=*SAS-data-set*

names an input data set that contains group summary statistics, decision limits, and outlier values in “strung out” form, with more than one observation per group. Each observation corresponds to one feature of one group’s box-and-whisker plot. Typically, this data set is created as an OUTBOX= data set in a previous run of the ANOM procedure with a BOXCHART statement. The BOX= data set is the only kind of summary data set you can use to produce schematic box-and-whisker plots. The BOXCHART statement is the only chart statement you can use with a BOX= input data set.

DATA=*SAS-data-set*

names an input data set that contains response values (typically, measurements or counts) as observations. Note that the DATA= data set may need sorting. If the values of the *group-variable* are numeric, you must sort the data set so that these

PROC ANOM and General Statements

values are in increasing order (within BY groups). Use PROC SORT if the data are not already sorted.

The DATA= data set may contain more than one observation for each value of the *group-variable*. This happens, for example, when you produce a chart for means and ranges with the XCHART statement.

You cannot use a DATA= data set together with a SUMMARY= or a TABLE= data set. If you do not specify one of these three input data sets, the ANOM procedure uses the most recently created data set as a DATA= data set. For more information, see the “DATA= Data Set” section in the chapter for the chart statement you are using.

GOUT=*graphics-catalog*

specifies the graphics catalog for graphics output from the ANOM procedure. This is useful if you want to save the output.

SUMMARY=*SAS-data-set*

names an input data set that contains group summary statistics. For example, you can read sample sizes, means, and standard deviations for the groups to create an ANOM chart. Typically, this data set is created as an OUTSUMMARY= data set in a previous run of the ANOM procedure, but it can also be created using a SAS summarization procedure such as PROC MEANS.

Note that the SUMMARY= data sets may need sorting. If the values of the *group-variable* are numeric, you need to sort the data set so that these values are in increasing order (within BY groups). Use PROC SORT if the data are not already sorted. The SUMMARY= data set can contain only one observation for each value for the *group-variable*.

You cannot use a SUMMARY= data set with a DATA= or a TABLE= data set. If you do not specify one of these three input data sets, the ANOM procedure uses the most recently created data set as a DATA= data set. For more information, see the “SUMMARY= Data Set” section in the chapter for the chart statement you are using.

LIMITS=*SAS-data-set*

names an input data set that contains preestablished decision limits or the parameters from which decision limits can be computed. Each observation in a LIMITS= data set provides decision limit information for a *response*. Typically, this data set is created as an OUTLIMITS= data set in a previous run of the ANOM procedure.

If you omit the LIMITS= option, then decision limits are computed from the data in the DATA= or SUMMARY= input data sets. For details about the variables needed in a LIMITS= data set, see the “LIMITS= Data Set” section in the chapter for the chart statement you are using.

TABLE=*SAS-data-set*

names an input data set that contains group summary statistics and decision limits. Each observation in a TABLE= data set provides information for a particular group and *response*. Typically, this data set is created as an OUTTABLE= data set in a previous run of the ANOM procedure.

Syntax for the PROC ANOM Statement

You cannot use a TABLE= data set with a DATA= or a SUMMARY= data set. If you do not specify one of these three input data sets, the ANOM procedure uses the most recently created data set as a DATA= data set. For more information, see the “TABLE= Data Set” section in the chapter for the chart statement that you are using.