



# Don't Be a SAS® Dinosaur: Modernize Your SAS Programs

by Warren Repole

## Frequency Reports: Number of Levels of a Variable

---

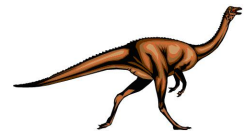
### Scenario:

You need to determine the number of distinct levels for several variables in a SAS data set.

If missing values are present, you wish to distinguish missing values (including regular and special missing values for numeric variables) from nonmissing values when the counts are generated.

---

### The old way: Count Levels from PROC FREQ Output



The original approach is to collapse each variable down to unique values, then count the number of collapsed items. The counts for individual variables subsequently are concatenated to create a single result set.

```
data Georgia;
  set sashelp.zipcode(where=(StateName="Georgia"));
  if msa=0 then msa=.;
run;
proc freq data=georgia;
  tables CountyNm / out=Counties noprint;
  tables AreaCode / out=AreaCodes noprint;
  tables ZIP_Class / out=ZIPclasses noprint;
  tables MSA / out=MSAs noprint;
run;
data counts;
  set Counties(in=county_level)
      AreaCodes(in=areacode_level)
      ZIPclasses(in=zipclass_level)
      MSAs(in=msa_level)
      end=no_more;
  if county_level then CountyCount+1;
  else if areacode_level then AreaCodeCount+1;
  else if zipclass_level then ZIPclassCount+1;
  else if msa_level then MSACount+1;
  if no_more then output;
  keep CountyCount AreaCodeCount ZIPclassCount MSACount;
run;
proc print data=counts noobs;
  title "Counts of Variable Levels";
run;
```

---

Please direct all correspondence to:

Warren Repole, 1705 Palm Springs Dr, Vienna VA 22182-2331

[sasdinosaur@repole.com](mailto:sasdinosaur@repole.com)

[www.repole.com/dinosaur](http://www.repole.com/dinosaur)

## The new way: Use the NLEVELS Option in PROC FREQ (available in SAS 9)

An alternate approach is to apply the NLEVELS option in PROC FREQ, using the ODS OUTPUT statement to capture the counts into a SAS data set.

```
data Georgia;
  set sashelp.zipcode(where=(StateName="Georgia"));
  if msa=0 then msa=.;
run;
ods output NLevels=CountsNLevels;
proc freq data=georgia nlevels;
  tables CountyNm AreaCode ZIP_Class MSA / noprint missing;
  label ZIP_Class=" " MSA=" " CountyNm=" " AreaCode=" ";
  title1 "Counts of Variable Levels";
run;
proc print data=CountsNLevels;
  title1 "Counts of Variable Levels";
run;
```

---

Advantages of the alternate approach:

- A single PROC step can analyze any number of variables.
- In addition to determining the number of unique values, levels of missing values are identified.
- If only a report is required, there is no need to create a data set to hold the results.

Disadvantages of the alternate approach:

- This approach produces a multiple-observation result set that must be transposed if one column per original variable is desired.

Additional documentation for this technique can be found in *Base SAS® 9.2 Procedures Guide*. Cary, NC: SAS Institute Inc.

Visit <http://support.sas.com/documentation/onlinedoc/sas9doc.html> for SAS 9 documentation.

Go to <http://www.repole.com/dinosaur/nlevels.html> for the sample code and output for this topic.

These techniques are mentioned in other SAS references and publications:

SAS Sample 30867 (<http://support.sas.com/kb/30/867.html>)

*What's New in PROC FREQ Procedure, Version 9*; Jyotheeswara Naidu Yellanki and Raj Suligavi,

<http://www.lexjansen.com/pharmasug/2005/coderscorner/cc23.pdf>.