STAT 541

Chapter 14: Using Advanced Character Functions

The LAG Function

- The LAG function retrieves a value of the specified variable from a previous observation.
- LAG2 will retrieve the value from two observations ago, etc.
- This can be useful for calculating moving averages for variables measured over time (see example).

COUNT and related functions

- The COUNT function works on character variables. It counts the number of times a specified substring appears in a character string.

- Similar functions: COUNTW counts the number of words in a character string.

- COUNTC counts the number of times any character from a specified list appears within a character string.

- If the modifier 'v' is included as an extra argument, then COUNTC returns the number of characters in the string that are NOT in the specified list.

- The modifier 'i' tells SAS to ignore case and treat uppercase and lowercase the same.

LENGTHN and LENGTHC functions

- The LENGTHN function returns the number of characters in a character string (not counting trailing blanks).

- If the character string is blank, LENGTHN returns 0.

- LENGTHC is similar, but includes trailing blanks if they occur.

FIND and Related Functions

- The FIND function returns the starting position of a specified substring within a character string.

- It is useful, along with the SUBSTR function, to locate and extract a portion of a string.

- FINDC is similar, but will return the starting position of any character (from a specified list of characters) within the string.

- FINDW will return the starting position of a specified word in a string.

- With the 'e' modifier, it will return which word in the string that the specified word is.

Pattern Matching

- Perl regular expressions allow you to process patterns in character strings and look for matches to specified patterns.

- These regular expressions use *metacharacters*, which are characters that have meanings can be represent text characters.

- For example, \d represents a digit (0-9) and \D represents a non-digit character.

- The start and end of a regular expression is delimited by /.

- Groups of characters are represented by ( ).

Pattern Functions

- PRXMATCH searches for a pattern match and returns the position where a pattern is found.

- If the pattern is not found is a data record, it returns a 0.

- PRXPARSE returns a pattern identifier number that can be used in other functions.

- PRXCHANGE is used to substitute text when a specified pattern is found.

- Its three arguments are: the Perl expression specifying the pattern; the number of times to replace it (-1 will replace all such patterns until the end of the source); and the source (a constant, a column, or an expression) that will be searched.