

That's the One: Choosing the Correct Statistical Test for Trend in SAS

OUTLINE

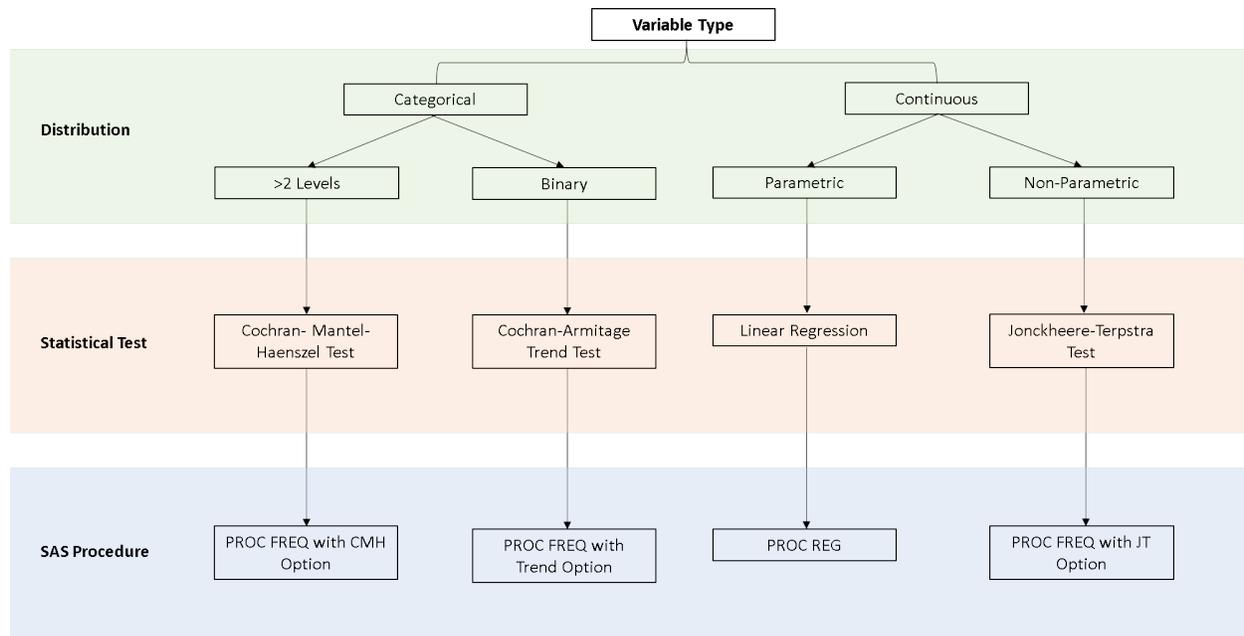
1. Abstract
2. Introduction
3. Which statistical test should I use to test for trend?
4. Which SAS procedure should I use to test for trend?
5. What do the results mean?
6. Considerations
7. Conclusion
8. References

ABSTRACT

Tests for trend are an informative and useful tool to examine the directionality of the differences in the means, medians or proportions of continuous or categorical variables across ordered groups. In clinical and epidemiological research, comparisons of baseline patient characteristics (e.g., demographic, clinical and laboratory data) across ordered levels of the categorized primary exposure are often examined with chi-square or analysis of variance (ANOVA) statistical tests, which identify the existence of differences but not trends in the ordered groups. Trend tests provide additional insight into the pattern of the relationship between independent and dependent variables. Multiple methods are available in SAS to evaluate trends of continuous and categorical variables using PROC REG (simple linear regression) and PROC FREQ (Cochran-Armitage, Jonckheere-Terpstra and Cochran-Mantel-Haenszel tests) statements. However, choosing the appropriate statistical test can be a challenge. Tests will vary depending on the assumptions about the variable of interest including its type and distribution. Selecting an inappropriate test may lead to incorrect inferences about the association of the variable across ordered exposure groups. This is important, especially when the results from trend tests may influence which variables are considered as covariates in models of adjustment. In this paper, we aim to (1) describe when to use specific statistical tests to evaluate trends in continuous or categorical variables across ordered groups, and (2) provide examples of SAS codes for trend tests and interpret the resulting output.

INTRODUCTION

Figure 1. Flowchart for Selecting the Appropriate Statistical Test for Trend and Corresponding SAS Procedure Depending on the Type and Distribution of the Variable of Interest.



Which statistical test should I use to test for trend?

*Will provide additional information for each test.

1. Continuous Variable – Normal distribution

Statistical Test: Simple Linear Regression

2. Continuous Variable – Non-normal distribution

Statistical Test: Jonckheere-Terpstra Test

3. Categorical Variable – Binary

Statistical Test: Cochran-Armitage Trend Test

4. Categorical Variable – More than 2 levels

Statistical Test: Cochran-Mantel-Haenszel Test

EXAMPLE: We will use data from the National Health and Nutrition Examination Survey (NHANES, 1999-2010) to demonstrate example code and output.

Which SAS procedure should I use to test for trend?

*Will provide description and distribution of variables and appropriate statistical tests.

Example 1: PROC REG

```
proc reg data=chol9c;
    model creage=hdl_cat;
run;
```

Example 2: PROC FREQ with <JT option>

```
proc freq data=chol9c;
    table LBXSTR*hdl_cat/noprint jt;
run;
```

Example 3: PROC FREQ with <Trend option>

```
proc freq data=chol9c;
    table sex*hdl_cat/trend norow nopercent;
run;
```

Example 4: PROC FREQ with <CMH option>

```
proc freq data=chol9c;
    table race*hdl_cat/cmh norow nopercent;
run;
```

What do the results mean?

*Will annotate output and provide interpretation of results.

Example 1

The SAS System					
The REG Procedure					
Model: MODEL1					
Dependent Variable: creage					
Number of Observations Read		31249			
Number of Observations Used		31249			
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2526.18903	2526.18903	6.52	0.0107
Error	31247	12106537	387.44637		
Corrected Total	31248	12109063			
Root MSE		19.68366	R-Square	0.0002	
Dependent Mean		47.56049	Adj R-Sq	0.0002	
Coeff Var		41.38657			
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	48.04224	0.21907	219.30	<.0001
hdl_cat	1	-0.13358	0.05231	-2.55	0.0107

Example 2

The SAS System

The FREQ Procedure

Statistics for Table of LBXSTR by hdl_cat

Jonckheere-Terpstra Test	
Statistic	204961346.0
Z	-42.6513
One-sided Pr < Z	<.0001
Two-sided Pr > Z	<.0001

Effective Sample Size = 31234
Frequency Missing = 15

Example 3

The SAS System

The FREQ Procedure

Frequency Col Pct	Table of sex by hdl_cat						
	hdl_cat						Total
	1	2	3	4	5	6	
sex							
1	9469 100.00	3141 100.00	2521 100.00	0 0.00	0 0.00	0 0.00	15131
2	0 0.00	0 0.00	0 0.00	1803 100.00	3717 100.00	10598 100.00	16118
Total	9469	3141	2521	1803	3717	10598	31249

Statistics for Table of sex by hdl_cat

Cochran-Armitage Trend Test	
Statistic (Z)	-166.2192
One-sided Pr < Z	<.0001
Two-sided Pr > Z	<.0001

Sample Size = 31249

Example 4**The SAS System****The FREQ Procedure**

Frequency Col Pct	Table of race by hdl_cat						
	race	hdl_cat					
		1	2	3	4	5	6
1	2199 23.22	619 19.71	410 16.26	480 26.62	999 26.88	2052 19.36	6759
2	619 6.54	167 5.32	131 5.20	155 8.60	307 8.26	684 6.45	2063
3	4753 50.20	1482 47.18	1131 44.86	810 44.93	1612 43.37	5166 48.75	14954
4	1506 15.90	741 23.59	757 30.03	290 16.08	651 17.51	2232 21.06	6177
5	392 4.14	132 4.20	92 3.65	68 3.77	148 3.98	464 4.38	1296
Total	9469	3141	2521	1803	3717	10598	31249

Summary Statistics for race by hdl_cat

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)				
Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	16.2961	<.0001
2	Row Mean Scores Differ	4	37.3001	<.0001
3	General Association	20	485.5525	<.0001

Total Sample Size = 31249

CONSIDERATIONS**CONCLUSION****REFERENCES**