

## Algorithm to Compare the Slopes (Regression Coefficients) between the subgroups in Simple/Multiple Regression using PROC REG

Sandeep Sawant, PharmaNet/i3, India, Mumbai

### ABSTRACT

Regression analysis is the most common technique used for data analysis in clinical trials. In regression analysis, a regression line is fitted for the response variable (e.g. Viral load at the end of the study) based on few explanatory variables (e.g. Baseline viral load, time since first diagnosis etc.). A regression line  $Y=a+bX$  is fitted, where Y is the response variable, X is the explanatory variable, a denotes the intercept and b is the slope (regression coefficient) of the line. The slope indicates the change in the value of Y if X is changed by one unit. Therefore slope is often useful measure of examining the rate of change in variable Y. In clinical trials, comparing slope (rate of change) for two (or more subgroups e.g. Active vs. Placebo) can be the area of interest to assess the effect of medical treatment. SAS® procedure PROC REG does not performs the desired analysis directly but some kind of data manipulation is needed. This paper will discuss the algorithm for comparing the regression coefficients for simple/multiple regression for 2 or more subgroups.

### INTRODUCTION

There are two types of regression

- a) Simple Regression ( $Y=a+bX$ )
- b) Multiple Regression ( $Y=a+bX+cZ+dW$ )

Let's discuss this situation by considering the following data. Suppose in a particular trial, there are 100 subjects receiving treatment XYZ. Subjects are randomized to receive 2 different doses (10mg, 50mg). The objective of the trial is to examine the optimum dose out of the 2 doses in terms of reduction in the viral load. A simple linear regression line can be fitted with viral load at the end of the study as response variable and baseline viral load as explanatory variables. The analysis can be performed at various dose levels. The regression coefficient for the two dose level will be the useful measure in determining the rate of change in viral load. The hypothesis of interest is to test the equality of the slopes for the 2 dose levels. If this hypothesis is accepted then we can conclude that there is no difference in the dose levels as such but if the hypothesis is rejected there can be evidence of superiority of the active treatment over the placebo.

Consider the following data

Patient (USUBJID)	Dose Level (DOSE)	Viral Load at EOS (RESP)	Viral Load at Baseline (EXP1)	Time since Diagnosis (EXP2)
100	10 mg	122	109	3.5
101	50mg	23	234	6.8
102	10 mg	124	45	7.2
103	50mg	90	309	4.1
104	10 mg	56	12	6.6
105	50mg	95	67	2.3
106	10 mg	88	99	42
107	50mg	67	134	45

In the above case, a simple linear regression line can be fitted as  $Y=a+bX$  for each dose level

Where Y=Viral load at the end of the study

X=Baseline viral load

a=Intercept

b=Regression Coefficient

We can interpret the regression coefficient as follows:

Positive regression coefficient indicates X and Y changes in same direction i.e. if X increases, Y increases. If X decreases, Y decreases. Negative regression coefficient indicates X and Y changes in opposite direction i.e. if X increases, Y decreases. If X decreases, Y increases.

If the treatment is effective, we would expect the negative regression coefficient. It would be also required to check if the rate of change in the viral load varies per the dose level and in this case it would be needed to check the hypothesis of equality of the slopes.

In this case,

Number of explanatory variable =1 (baseline viral load)

Number of subgroups=2 (10mg, 50mg)

If we add another explanatory variable, time since diagnosis for the patient in the above simple regression, this would be the case of multiple regression model with

Number of explanatory variable =2 (baseline viral load and time since diagnosis)

Number of subgroups=2 (10mg, 50mg)

The desired regression statistics can be summarized in following table

Parameter	Treatment Group			
	XYZ 10mg (N=XXX)		XYZ 50mg (N=XXX)	
Rate of Change (Slope)	Estimate xxx	95% CI (xx.x, xx.x)	Estimate xxx	95% CI (xx.x, xx.x)
Treatment Comparison	Difference	xx.x		
	95% CI	(xx.x, xx.x)		
	p-value	0.xxxx		

The above table displays regression coefficients (Slopes) obtained from each of the regression line. The treatment comparison and p-value from the table can be used to test the hypothesis of equality of rate of change in viral load between the two dose levels.

### **Algorithm to derive the regression Statistics:**

#### **Step 1:**

Fit the Regression Model with the desired response and exploratory variables.

```
proc reg ;
  by TRT;
  model RESP= EXP1 EXP2/clb;
  ods output parameterestimates=param;
run;
quit;
```

The PARAM dataset will contain the slopes and corresponding 95% CI at each dose level.

#### **Step 2:**

Identify the number of subgroups (denoted by m) and the number of exploratory variables (denoted by n).

In this case m=2 as there are two dose levels (10mg, 50mg)

In this case n=2 (EXP1 and EXP2).

### **Step 3:**

Create (m-1) dummy variables as dummy1, dummy2 etc.

In this case, we create dummy1 as

Dummy1=1 if TRT='10 mg'  
=0 if TRT='^10 mg'

If there are 3 subgroups (10mg, 50mg and 100mg), create two dummy variables as

Dummy1=1 if TRT='10 mg'  
=0 if TRT='^10 mg'

Dummy2=1 if TRT='50 mg'  
=0 if TRT='^50 mg'

### **Step 4:**

Create the cross product for each of the dummy variable and exploratory variable.

In this case we create

Expdum1=EXP1\*Dummy1

Expdum2=EXP2\*Dummy1

In general, create (m-1)\*n cross product terms.

### **Step 5:**

Fit the final regression analysis model with the dummy and cross product terms as follows

```
proc reg ;  
  model RESP= EXP1 EXP2 EXPDUM1 EXPDUM2 DUMMY1/clb;  
  ods output parameterestimates=param1;  
run;  
quit;
```

The parameter estimates against EXPDUM1 (or EXPDUM2 whichever is needed on the table) will provide the difference in slope and corresponding 95% CI. The corresponding p-value is useful in testing the hypothesis of equality of the slopes.

## **CONCLUSION**

The above algorithm can be used to get difference in slopes and to test the hypothesis of equality of the slopes. The algorithm is useful in simple or multiple regression with 2 or more subgroups.

## **ACKNOWLEDGMENTS**

I thank all my colleagues for guiding this work and carefully reviewing the paper with comments and suggestions.

## **CONTACT INFORMATION**

Your comments and questions are valued and encouraged. Contact the author at:

Sandeep Sawant

PharmaNet/i3, 7<sup>th</sup> Floor, Corporate Center, Opp. To VITS Hotel,

Andheri-Kurla Road, Andheri (E)- 400059, Mumbai, India

Work Phone: +91-22-30554032

E- mail: [sandeep.sawant@i3global.com](mailto:sandeep.sawant@i3global.com)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.