

Open Problem for SUAVE Meeting November 27th, at UVic

The open problem for the upcoming SUAVE meeting is to write a short SAS program that calculates R^2 based on a table of sums of squares that has previously been determined.

Use the following small 5 row dataset (which includes a row where all values are missing):

Source	DF	SS	MS
Model	2	6197.1	3098.6
Error	62	2775.9	44.8
Uncorrected Total	64	8973.0	—
	—	—	—
Corrected Total	63	3257.7	—

This data depicts the typical analysis of variance (ANOVA) table that results after running proc nlin (nonlinear regression) with the 'totalss' option (for those that are interested, the ODS table is named 'anova'). This table conveys how the data's variability (i.e. both the sums of squares (SS) and mean squares (MS)) has been partitioned.

A popular measure of goodness-of-fit for non-linear models is the pseudo coefficient of determination or pseudo- R^2 , which can be calculated using: $R^2 = 1 - \frac{SS(Error)}{SS(Corrected\ Total)}$.

That is, R^2 is the compliment of the ratio of two sums of squares. In the ratio, the numerator is the SS associated with the Error and the denominator is the SS associated with the Corrected Total.

Your task, should you choose to accept it, is to write a short SAS program that calculates R^2 using the above small dataset. The program must run successfully without **any** warnings or errors in the log. Bonus marks will be given to efficient/compact coding.

Good luck! Please let one of the SUAVE executive know by November 20th if you'd like to share your solution at the SUAVE meeting!

See below for solutions...

Solution #1. Gord Nigh (BC Ministry of Forests, Lands and Natural Resource Operations)

```
data anovatable;
  format source $8.;
  retain sse;
  input source $ @;
  if source = 'Error' then do;
    input df sse mse;
    delete;
  end;
  else if source = 'Correcte' then do;
    input junk $ df ssct;
    Rsquared = 1 - (sse / ssct);
  end;
  else delete;
  label sse='Sum of#squared#errors' ssct='Corrected#total#sum of#squared#errors' Rsquared='R-squared';
cards;
Source      DF      SS      MS
Model 2      6197.1 3098.6
Error 62      2775.9 44.8
Uncorrected Total 64      8973.0
Corrected Total 63      3257.7
;
run;

proc print split='#' noobs; var sse ssct Rsquared;
run;

quit;
```

Solution #2. Mike Atkinson (Acko Systems Consulting Inc.)

```
* read the data;
data totalss;
  infile cards4;
  length source $ 20;
  input source $ df ss ms;
  source = translate(source, ' ', '_');
cards4;
Model 2      6197.1 3098.6
Error 62      2775.9 44.8
Uncorrected_Total 64      8973.0
Corrected_Total 63      3257.7
;;;
run;

* to print the results;
proc sql;
  select max(case when upcase(source) = 'ERROR' then ss else . end) as ss_error,
         max(case when upcase(source) = 'CORRECTED TOTAL' then ss else . end) as ss_corrected,
         1 - (calculated ss_error / calculated ss_corrected) as r_squared
  from totalss;
quit;

* OR, to put the value into a macro variable;
proc sql noprint;
  select max(case when upcase(source) = 'ERROR' then ss else . end) as ss_error,
         max(case when upcase(source) = 'CORRECTED TOTAL' then ss else . end) as ss_corrected,
         1 - (calculated ss_error / calculated ss_corrected) as r_squared
  into :ss_e, :ss_c, :r_squared
  from totalss;
quit;
```

```
%put R-Squared = &r_squared;
```

Solution #3. Jim Goudie (BC Ministry of Forests, Lands and Natural Resource Operations)

```
/*
```

```
Note: I assumed the "-" values in the ANOVA tables were missing values because they are treated as  
such if the table is output to a data set using an "ODS output" statement in PROC NLIN.
```

```
*/
```

```
data Anova;
```

```
input Source $char17. DF 21-22 SS 24-30 MS 32-37;
```

```
if source ne "";*<-----eliminates null line;
```

```
datalines;
```

```
Model 2 6197.1 3098.6
```

```
Error 62 2775.9 44.8
```

```
Uncorrected Total 64 8973.0
```

```
Corrected Total 63 3257.7
```

```
;
```

```
run;
```

```
proc transpose data=Anova out=trans_anova;
```

```
var ss;
```

```
id source;
```

```
idlabel source;
```

```
run;
```

```
data trans_anova;
```

```
set trans_anova ;
```

```
R_SQ=1-(Error/Corrected_Total);*<-----calculate pseudo r-squared statistic;
```

```
run;
```

```
proc print;
```

```
run;
```

```
quit;
```