

3.3: Anatomy of SAS Programming for ANOVA

The statistical software SAS is widely used in this course, and in previous sections we came across outputs generated through SAS programs. In this section, we begin to delve further into SAS programming with a special focus on ANOVA-related statistical procedures. The STAT 480-course series is also a useful resource for additional help.

Here is the program used to generate the summary output in Section 2.1:

```
data greenhouse;  
input Fert $ Height;
```

The first line begins with the word `data` and invokes the data step. Notice that the end of each SAS statement has a semicolon. This is essential. In the dataset, the data to be used and its variables are named. Note that SAS assumes variables are numeric in the input statement, so if we are going to use a variable with alpha-numeric values (e.g. F1 or Control), then we have to follow the name of the variable in the input statement with a `$` sign.

A simple way to input small datasets is shown in this code, wherein we embed the data in the program. This is done with the word `datalines`.

```
datalines;  
Control    21  
Control    19.5  
Control    22.5  
Control    21.5  
Control    20.5  
Control    21  
F1         32  
F1         30.5  
F1         25  
F1         27.5  
F1         28  
F1         28.6  
F2         22.5  
F2         26  
F2         28  
F2         27  
F2         26.5  
F2         25.2  
F3         28  
F3         27.5  
F3         31  
F3         29.5  
F3         30  
F3         29.2  
;
```

The semicolon here ends the dataset.

SAS then produces an output of interest using `proc` statements, short for “procedure”. You only need to use the first four letters, so SAS code is full of `proc` statements to do various tasks. Here we just wanted to print the data to be sure it read it in OK.

```
proc print data= greenhouse;
title 'Raw Data for Greenhouse Data'; run;
```

Notice that the data set to be printed is specified in the `proc print` command. This is an important habit to develop because if not specified, SAS will use the last created data set, out of both input data sets, and output datasets that may have been generated as a result of any SAS procedures run up to that point.

The summary procedure which was then run can be very useful in both EDA (exploratory data analysis) and obtaining descriptive statistics such as mean, variance, minimum, maximum, etc. SAS procedures including the summary procedure categorical variables are specified in the class statement. Any variable NOT listed in the class statement is treated as a continuous variable. The target variable for which the summary will be made is specified by the `var` (for variable) statement.

The `output` statement creates an output dataset and the `out=` part assigns a name of your choice to the output. Descriptive statistics also can be named. For example, in the `output` statement below, `mean=mean` and `stderr=se` have named the mean of the variable `fert` as `mean` and standard error as `se`. The output data sets of any SAS procedure will not be automatically printed. As illustrated in the code below, the print procedure would then have to be used to print the generated output. In the `proc print` command a title can be included as a means of identifying and describing the output contents.

```
proc summary data= greenhouse;
class fert;
var height;
output out=output1 mean=mean stderr=se;
run;
proc print data=output1;
title 'Summary Output for Greenhouse Data';
run;
```

The two commands `title ; run;` right after will erase the title assignment. This prevents the same title to be used in every output generated thereafter, which is a default feature in SAS.

```
title; run;
```

Summary Output for Greenhouse Data

Obs	Fert	TYPE	FREQ	mean	se
1			0	24	26.1667
2	Control		1	6	21.0000
3	F1		1	6	28.6000
4	F2		1	6	25.8667
5	F3		1	6	29.2000

This page titled [3.3: Anatomy of SAS Programming for ANOVA](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Penn State's Department of Statistics](#).