

# Fast Facts for SAS

## Biostat 510

1. Read in **raw data** from an ASCII file using an **infile** statement.

```
data march;
  infile "marflt.dat";
  input   flight 1-3
         @4 date mmddy6.
         @10 time time5.
         orig $ 15-17
         dest $ 18-20
         @21 miles comma5.
         mail 26-29
         freight 30-33
         boarded 34-36
         transfer 37-39
         nonrev 40-42
         deplane 43-45
         capacity 46-48;
  format date mmddy10. time time5. miles comma5.;
  label flight="Flight number"
         orig  ="Origination City"
         dest  ="Destination City";
run;
```

2. Import an **Excel File** using **Proc Import** (alternatively, use the **Import Wizard**):

```
PROC IMPORT OUT= WORK.MARCH
            DATAFILE= "MARCH.XLS"
            DBMS=EXCEL REPLACE;
  SHEET="march$";
  GETNAMES=YES;
  MIXED=NO;
  SCANTEXT=YES;
  USEDATE=YES;
  SCANTIME=YES;
RUN;
```

3. Read in **raw data** from a **CSV** (comma separated values) file.

```
data pulse;
  infile "pulse.csv" firstobs=2 delimiter="," missover;
  input pulse1 pulse2 ran smokes sex height weight activity;
run;
```

4. Convert an **SPSS portable file** into a SAS data set:

```
filename file1 "cars.por";  
proc convert spss=file1 out=cars;  
run;
```

5. Read in a **Permanent SAS data set**, and create a temporary data set:

```
libname sasdata2 "C:\Documents and Settings\kwelch\Desktop\sasdata2";  
data bank;  
  set sasdata2.bank;  
run;
```

Or, to **use the permanent SAS data set** for analysis directly:

```
libname sasdata2 "C:\Documents and Settings\kwelch\Desktop\sasdata2";  
proc means data=sasdata2.bank;  
run;
```

Another way to **use the permanent SAS data set** directly, without setting up a **libname** statement:

```
proc means data="C:\Documents and Settings\kwelch\Desktop\sasdata2\bank.sas7bdat";  
run;
```

Or:

```
proc means data="c:\Documents and Settings\kwelch\Desktop\sasdata2\bank";  
run;
```

6. Read a SAS transport file into a regular SAS data set:

```
libname trans xport "c:\temp\sasdata2\bank.xpt";  
proc copy in=trans out=sasdata2;  
run;
```

#### 7. Rules for SAS statements:

- They start with a keyword, such as **proc** or **var**.
- They can be any length.
- They end with a semicolon (;).

#### 8. Rules for SAS names:

- They can have only letters, numbers, and underscores in them.
- They may not start with a number.
- They may not have any blanks.
- They can be upper or lower case.
- SAS versions 7 through 9 allow variable names of up to 32 characters.
- SAS version 6 only allows variable names of up to 8 characters.
- Library names must only be 8 characters or less.

#### 9. SAS Data step:

- Used for creating or modifying a data set, adding new variables.
- Start with a **data** statement.
- End with a **run** statement.
- Statements are (usually) processed in order from top to bottom.
- Data step usually does not produce any output in output window.
- Check **log** to be sure data set was created properly.

#### 10. SAS Proc step:

- Used for analysis or generating a report.
- Start with a **proc** statement.
- Often, but not always, **produce output** in the output window.
- End with a **run** statement, or a **run** statement and **quit** statement.

#### 11. Procs for working with Categorical Data:

Descriptives:

Proc Freq (numeric or character variables)  
 Single variable: oneway tabulation  
 Two or more variables: crosstabs

Simple Statistical Tests:

One variable (with 2 or more levels)  
 Proc Freq (chi-square goodness of fit test)  
 Two variables (each with 2 or more levels)  
 Proc Freq (chi-square test of independence)

Graphs:

Proc Gchart (bar charts)

Modeling:

Proc Logistic  
 Logistic regression models for binary or ordinal dependent variables  
 Proc Genmod  
 Generalized linear models for count, binary, or other dependent variables (exponential family of distributions); predictors may be nominal, ordinal, or continuous.  
 Proc Glimmix  
 Generalized linear mixed models for count or binary dependent variable, longitudinal or clustered data (exponential family); predictors may be nominal, ordinal, or continuous.

#### 12. Procs for working with Continuous data:

Descriptives:

Proc Means  
 Proc Univariate

Graphs:

Proc Univariate (histograms, qqplots)  
 Proc Gplot (bivariate scatter plots, regression plots)  
 Proc Boxplot (plot continuous vs. categorical, sort first)

Simple statistical tests:

One Sample  
 Proc Univariate (one-sample t-test, nonparametric tests)  
 Proc ttest (one-sample t-test)