Evaluating the model’s predictive ability

- borrow terminology from screening tests (tests that can be used to diagnose disease before any observable symptoms are present)
- characteristics of the test are evaluated by giving the test to people whose disease status is known
- then test is used to try to determine disease status of people who are not yet showing symptoms
- idea extends to logistic regression
  - the model gives predictions of the probability that each subject in the dataset has a “yes” for the response variable
  - we intend to use the model to predict the response variable status for future individuals for whom values of explanatory variables are known but value of response variable is not yet known

Four aspects of predictive ability

- sensitivity
  - the probability that screening test is positive given that the person has the disease
- specificity
  - the probability that a symptom is not present (or screening test is negative) given that the person does not have the disease
- positive predictive value
  - the probability that a person has the disease given a positive test result
- negative predictive value
  - the probability that a person does not have the disease given a negative test
Iterative and conditional processing in SAS macro programs

- **mprint** option was used so that SAS log will show code actually generated by macro programs.

```sas
%macro daily;
  proc means data=books.ytdsales(where=(datesold=today()))
    maxdec=2 sum;
  title "Daily Sales Report for &sysdate";
  class section;
  var salepric;
  run;
  %if &sysday=Friday %then %do;
    proc means data=books.ytdsales
      (where=(today()-6 le datesold le today()))
    sum maxdec=2;
    title "Weekly Sales Report Week Ending &sysdate";
    class section;
    var salepric;
    run;
  %end;
%mend daily;

%daily
```  

Some useful character functions in SAS

- Character functions operate on the values of character variables.
- Also called string functions.
The Compress function: removing selected characters

- **compress** function with single argument removes only blanks
- optional second argument is list of all characters to be removed (enclosed in single or double quotes)

```sas
options linesize = 72 ;

data phonebook ;
length first last $ 10 ;
input first last phone $ 21-33 ;
phone1 = compress(phone) ;
phone2 = compress(phone, '()- ') ;
datalines ;
Kate Cowles (319)354-3684
Brendan Holly 3193543684
Mysterious Stranger 515 555 1212 ;
proc print data = phonebook ;
run ;
```

**SUBSTR:** extract a part of a character variable

- arguments
  - character variable
  - which position to start at
  - how many characters to extract
- extracted sections will be padded out with blanks to length of original variable

```sas
data phonebook ;
set phonebook ;
area = substr(phone2,1,3) ;
exchg = substr(phone2,4,3) ;
rest = substr(phone2,7,4) ;
phone3 = '(' || area || ')' || exchg || '-' || rest ;
run ;
```

The concatenation operator: `||`

- glues together pieces to make a complete character string

```sas
proc print data = phonebook ;
run ;
```

```sas
Obs first last phone phone1 phone2
1 Kate Cowles (319)354-3684 (319)354-3684 3193543684
2 Brendan Holly 3193543684 3193543684 3193543684
3 Mysterious Stranger 515 555 1212 5155551212 5155551212
```

```sas
Obs area exchg rest phone3
1 319 354 3684 (319 )354 -3684
2 319 354 3684 (319 )354 -3684
3 515 555 1212 (515 )555 -1212
```
Fixed version

data phonebook;
set phonebook;
area = substr(phone2,1,3);
exchg = substr(phone2,4,3);
rest = substr(phone2,7,4);
phone3 = '(' || compress(area) || ')' || compress(exchg) || '-
' || compress(rest);
run;

proc print data = phonebook;
run;

Contents of phonebook dataset

The CONTENTS Procedure

Data Set Name: WORK.PHONEBOOK Observations: 3
Member Type: DATA Variables: 9
Engine: V8 Indexes: 0
Created: 10:43 Wednesday, July 16, 2003 Observation Length: 140
Last Modified: 10:43 Wednesday, July 16, 2003 Deleted Observations: 0
Protection: Compressed: NO
Data Set Type: Sorted: NO
Label:

-----Engine/Host Dependent Information-----

Data Set Page Size: 16384
Number of Data Set Pages: 1
First Data Page: 1
Max Obs per Page: 116
Obs in First Data Page: 3
Number of Data Set Repairs: 0
File Name: /usr/tmp/SAS_workF4D500005960_ mouse/phonebook.sas7bdat
Release Created: 8.0202M0
Host Created: HP-UX
Inode Number: 53106
Access Permission: rw-------
Owner Name: UNKNOWN
File Size (bytes): 24576

-----Alphabetic List of Variables and Attributes-----

# Variable Type Len Pos
-----------------------------------
 6   area   Char  13  59
 7  exchg   Char  13  72
 1  first   Char  10  0
 2  last    Char  10  10
 3  phone   Char  13  20
 4  phone1  Char  13  33
 5  phone2  Char  13  46
 9  phone3  Char  42  98
 8  rest    Char  13  85