Descriptive

codebook chd

```
-------------------------------------------------------------------------------
chd                                          Evidence of Coronary Heart Disease
---------------------------------------------------------------------------
                  type:  numeric (double)
                 label:  CHD

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/100

            tabulation:  Freq.   Numeric  Label
                            57         0  No
                            43         1  Yes
```

tab chd

```
Evidence of |
   Coronary |
      Heart |
    Disease |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |         57       57.00       57.00
        Yes |         43       43.00      100.00
------------+-----------------------------------
      Total |        100      100.00
```

Logistic

logistic chd age

```
Logistic regression                               Number of obs   =        100
                                                  LR chi2(1)      =      29.31
                                                  Prob > chi2     =     0.0000
Log likelihood = -53.676546                       Pseudo R2       =     0.2145

------------------------------------------------------------------------------
         chd | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   1.117307   .0268822     4.61   0.000     1.065842    1.171257
------------------------------------------------------------------------------
```

listcoef

listcoef, percent help

```
logistic (N=100): Percentage Change in Odds

  Odds of: Yes vs No

----------------------------------------------------------------------
         chd |      b         z     P>|z|      %      %StdX      SDofX
-------------+--------------------------------------------------------
         age |   0.11092    4.610   0.000     11.7    267.0    11.7213
----------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
       % = percent change in odds for unit increase in X
   %StdX = percent change in odds for SD increase in X
   SDofX = standard deviation of X
```

prchange

```
prchange age, help

logistic: Changes in Probabilities for chd

     min->max      0->1     -+1/2    -+sd/2  MargEfct
age    0.8690    0.0006    0.0267    0.3037    0.0267

              No      Yes
Pr(y|x)  0.5955  0.4045

             age
    x=     44.38
sd(x)=   11.7213

 Pr(y|x): probability of observing each y for specified x values
Avg|Chg|: average of absolute value of the change across categories
Min->Max: change in predicted probability as x changes from its minimum to
          its maximum
    0->1: change in predicted probability as x changes from 0 to 1
   -+1/2: change in predicted probability as x changes from 1/2 unit below
          base value to 1/2 unit above
  -+sd/2: change in predicted probability as x changes from 1/2 standard
          dev below base to 1/2 standard dev above
MargEfct: the partial derivative of the predicted probability/rate with
          respect to a given independent variable
```

```
prchange age, delta(10)

logistic: Changes in Probabilities for chd

(Note: d = 10)

     min->max      0->1     -+d/2    -+sd/2  MargEfct
age    0.8690    0.0006    0.2612    0.3037    0.0267

              No      Yes
Pr(y|x)  0.5955  0.4045   <---------------------    This is the probability for
                                                    CHD = "Yes" at mean of age
             age
    x=     44.38
sd(x)=   11.7213
```

prvalue

```
prvalue, x(age=44.38)

logistic: Predictions for chd

Confidence intervals by delta method

                                      95% Conf. Interval
  Pr(y=Yes|x):         0.4045    [ 0.2913,    0.5177]
  Pr(y=No|x):          0.5955    [ 0.4823,    0.7087]

      age
x=   44.38
```

Classification table

estat

```
estat class, cutoff(0.501)
```

```
Logistic model for chd

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |        29            12  |         41
     -     |        14            45  |         59
-----------+--------------------------+-----------
   Total   |        43            57  |        100

Classified + if predicted Pr(D) >= .501
True D defined as chd != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   67.44%
Specificity                     Pr( -|~D)   78.95%
Positive predictive value       Pr( D| +)   70.73%
Negative predictive value       Pr(~D| -)   76.27%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   21.05%
False - rate for true D         Pr( -| D)   32.56%
False + rate for classified +   Pr(~D| +)   29.27%
False - rate for classified -   Pr( D| -)   23.73%
--------------------------------------------------
Correctly classified                        74.00%
--------------------------------------------------
```

Same order as in SPSS:

```
predict pr
```

```
gen predicted_cut05 = pr > 0.5
```

```
tab2 chd  predicted_cut05, row
```

```
  Evidence |
        of |
  Coronary |
     Heart |    predicted_cut05
   Disease |         0          1 |     Total
-----------+----------------------+----------
        No |        45         12 |        57
           |     78.95      21.05 |    100.00
-----------+----------------------+----------
       Yes |        14         29 |        43
           |     32.56      67.44 |    100.00
-----------+----------------------+----------
     Total |        59         41 |       100
           |     59.00      41.00 |    100.00
```

```
display (45+29)/100
.74
SPSS was:
```

Classification Table[a]

	Observed		Predicted: Evidence of Coronary Heart Disease No	Predicted: Evidence of Coronary Heart Disease Yes	Predicted: Percentage Correct
Step 1	Evidence of Coronary Heart Disease	No	45	12	78,9
		Yes	14	29	67,4
	Overall Percentage				74,0

a. The cut value is ,500

prtab

```
prtab age

logistic: Predicted probabilities of positive outcome for chd

----------------------
Age       |
(years)   | Prediction
----------+-----------
       20 |     0.0435
       23 |     0.0596
       24 |     0.0662
       25 |     0.0733
       26 |     0.0812
       28 |     0.0994
       29 |     0.1098
…
       59 |     0.7747
       60 |     0.7934
       61 |     0.8110
       62 |     0.8274
       63 |     0.8427
       64 |     0.8569
       65 |     0.8699
       69 |     0.9125
----------------------
      age
x=  44.38
```

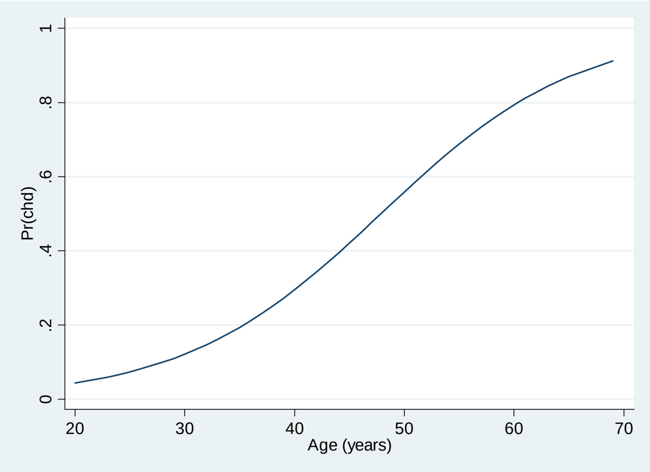
Test differences

```
prvalue, x(age=30) save

logistic: Predictions for chd

Confidence intervals by delta method

                                95% Conf. Interval
  Pr(y=Yes|x):        0.1211   [ 0.0265,    0.2157]
  Pr(y=No|x):         0.8789   [ 0.7843,    0.9735]

    age
x=   30

rvalue, x(age=60)  dif

logistic: Change in Predictions for chd

Confidence intervals by delta method

                     Current     Saved     Change   95% CI for Change
  Pr(y=Yes|x):        0.7934    0.1211     0.6723  [ 0.4822,    0.8625]
  Pr(y=No|x):         0.2066    0.8789    -0.6723  [-0.8625,   -0.4822]

          age
Current=   60
  Saved=   30
   Diff=   30
```

Plots

postgr3

```
postgr3 age
```

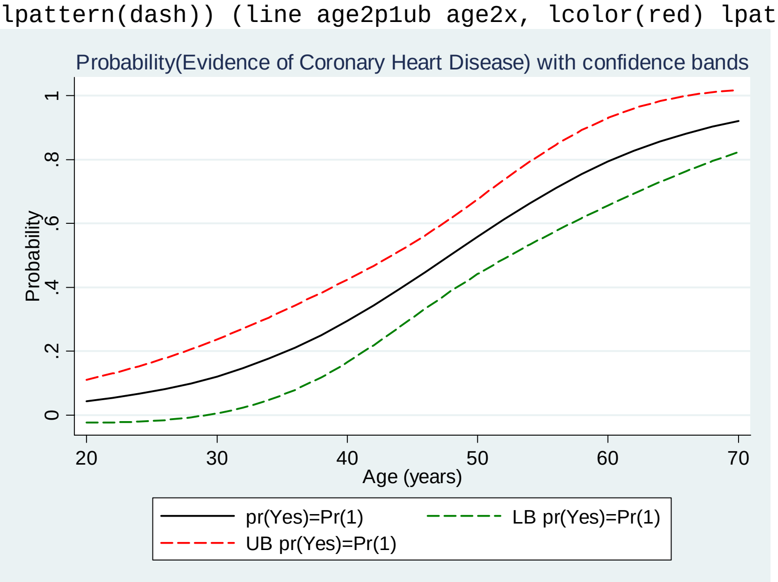


Plot with confidence

```
prgen age, from (20) to (70) gap(2) generate (age2) rest(mean) ci

* creates new variables:
```

Variable	Label (automatic)	explanation
age2x	Age (years)	takes values from 20 to 70 by 2
age2p0	pr(No)=Pr(0)	prob(0)
age2p1	pr(Yes)=Pr(1)	prob(1)
age2p0lb	LB pr(No)=Pr(0)	low bound of ci 95% for Pr(0)
age2p1lb	LB pr(Yes)=Pr(1)	low bound of ci 95% for Pr(1)
age2p0ub	UB pr(No)=Pr(0)	upper bound of ci 95% for Pr(0)
age2p1ub	UB pr(Yes)=Pr(1)	upper bound of ci 95% for Pr(1)

```
twoway (line age2p1 age2x, lcolor(black)) (line age2p1lb age2x, lcolor(green)
lpattern(dash)) (line age2p1ub age2x, lcolor(red) lpattern(dash))
```



Standardized residuals and leverage

Leverage = dbeta = Pregibon Delta-Beta influence statistic = counterpart to Cook'distance in linear regression

"We have seen quite a few logistic regression diagnostic statistics. Now how large does each one have to be, to be considered influential?

First of all, we always have to make our judgment based on our theory and our analysis.
Secondly, there are some rule-of-thumb cutoffs when the sample size is large. These are shown below.

When the sample size is large, the asymptotic distribution of some of the measures would follow some standard distribution.
That is why we have these cutoff values, and why they only apply when the sample size is large enough.

Usually, we would look at the relative magnitude of a statistic an observation has compared to others.

That is, we look for data points that are farther away from most of the data points. "

Measure	Value
leverage (hat value)	>2 or 3 times of the average of leverage
abs(Pearson Residuals)	> 2
abs(Deviance Residuals)	> 2

http://www.ats.ucla.edu/stat/Stata/webbooks/logistic/chapter3/statalog3.htm

Standardized residuals

```
predict prob
label var prob  "Prob(CHD=1)"
```

```
predict stdres, res
label var std "Std residual"

gen predict="yes" if prob>0.5
replace predict="no" if prob<=0.5
```

```
list  stdres age if abs(stdres)>2
```

```
     +----------------+
     |   stdres   age |
     |----------------|
  4. | 2.314473    25 |
  5. | 2.314473    25 |
     +----------------+
```

Leverage

```
predict cook, dbeta
label var cook "Cook's statistics"

gen obs=_n
scatter  cook obs, mlabel(obs) msize(1) mlabsize(2) jitter(*10)
```