```
      name:  <unnamed>
       log:  C:\Users\USER\Downloads\heart disease\multivariate logistics.smcl
  log type:  smcl
 opened on:  15 Jan 2026, 10:24:59

1 . *****multivariable logistic regression.

2 . *****describe each variable to check which are numeric and which are categoric
  > al

3 . describe gender

  Variable      Storage   Display    Value
      name         type    format    label      Variable label
  ------------------------------------------------------------------------------
  gender          str6    %9s                  gender: "Male", "Female" or
                                                 "Other"

4 . ****transform gender to numeric

5 . encode gender, generate (gender_num)

6 . describe gender_num

  Variable      Storage   Display    Value
      name         type    format    label      Variable label
  ------------------------------------------------------------------------------
  gender_num      long    %8.0g      gender_num
                                               gender: "Male", "Female" or
                                                 "Other"

7 . describe age

  Variable      Storage   Display    Value
      name         type    format    label      Variable label
  ------------------------------------------------------------------------------
  age             float   %9.0g

8 . describe hypertension

  Variable      Storage   Display    Value
      name         type    format    label      Variable label
  ------------------------------------------------------------------------------
  hypertension    byte    %8.0g                0 if the patient doesn't have
                                                 hypertension, 1 if the patient
                                                 has hypertension

9 . describe heart_disease

  Variable      Storage   Display    Value
      name         type    format    label      Variable label
  ------------------------------------------------------------------------------
  heart_disease   byte    %8.0g                0 if the patient doesn't have any
                                                 heart diseases, 1 if the patient
                                                 has a heart d
```

```
10 . describe avg_glucose_level

Variable      Storage   Display    Value
    name         type    format    label      Variable label
------------------------------------------------------------------------------
avg_glucose_l~l float   %9.0g                 average glucose level in blood

11 . describe bmi

Variable      Storage   Display    Value
    name         type    format    label      Variable label
------------------------------------------------------------------------------
bmi             float   %9.0g                 body mass index

12 . ****now that the relevant variables have been inspected and transformed, logis
> tic regression can be performed

13 . logistic stroke i.smoke_num i.gender_num bmi avg_glucose_level i.heart_disease
>  i.hypertension age
note: 3.gender_num != 0 predicts failure perfectly;
      3.gender_num omitted and 1 obs not used.


Logistic regression                                     Number of obs =  4,908
                                                        LR chi2(9)    = 358.92
                                                        Prob > chi2   = 0.0000
Log likelihood = -684.69076                             Pseudo R2     = 0.2077

------------------------------------------------------------------------------
      stroke | Odds ratio   Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
    smoke_num |
formerly s.. |   1.293017   .3171244     1.05   0.295     .7995377    2.091073
never smoked |   1.220409   .2813119     0.86   0.388     .7767803      1.9174
      smokes |   1.782901   .4691578     2.20   0.028     1.064485    2.986173
             |
   gender_num |
        Male |   .9888679   .1522966    -0.07   0.942     .7312128    1.337312
       Other |          1  (empty)
             |
         bmi |   1.003464   .0117854     0.29   0.768     .9806287    1.026831
avg_glucose~l |   1.004708    .001295     3.64   0.000     1.002173    1.007249
1.heart_dis~e |   1.451846   .2991842     1.81   0.070     .9694207    2.174348
1.hypertens~n |   1.678077    .292721     2.97   0.003     1.192147    2.362078
         age |   1.071481   .0062651    11.81   0.000     1.059271    1.083831
       _cons |   .0003094    .000181   -13.81   0.000     .0000983    .0009738
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.

14 . tabulate gender_num

    gender:
    "Male",
"Female" or
    "Other" |      Freq.     Percent        Cum.
------------+-----------------------------------
     Female |      2,994       58.59       58.59
       Male |      2,115       41.39       99.98
      Other |          1        0.02      100.00
------------+-----------------------------------
      Total |      5,110      100.00
```

```
15 . logistic stroke ib3.smoke_num i.gender_num bmi avg_glucose_level i.heart_disea
   > se i.hypertension age
   note: 3.gender_num != 0 predicts failure perfectly;
         3.gender_num omitted and 1 obs not used.


   Logistic regression                                     Number of obs =  4,908
                                                           LR chi2(9)    = 358.92
                                                           Prob > chi2   = 0.0000
   Log likelihood = -684.69076                             Pseudo R2     = 0.2077

   ------------------------------------------------------------------------------
         stroke | Odds ratio   Std. err.      z    P>|z|     [95% conf. interval]
   -------------+----------------------------------------------------------------
       smoke_num |
        Unknown |   .8193973   .1888762    -0.86   0.388     .5215396    1.287365
   formerly s.. |   1.059494   .1991555     0.31   0.759     .7329871    1.531443
         smokes |   1.460904    .311378     1.78   0.075     .9620467    2.218438
                |
      gender_num |
           Male |   .9888679   .1522966    -0.07   0.942     .7312128    1.337312
          Other |          1  (empty)
                |
            bmi |   1.003464   .0117854     0.29   0.768     .9806287    1.026831
   avg_glucose~l |   1.004708    .001295     3.64   0.000     1.002173    1.007249
   1.heart_dis~e |   1.451846   .2991842     1.81   0.070     .9694207    2.174348
   1.hypertens~n |   1.678077    .292721     2.97   0.003     1.192147    2.362078
             age |   1.071481   .0062651    11.81   0.000     1.059271    1.083831
           _cons |   .0003776   .0002157   -13.80   0.000     .0001233    .0011567
   ------------------------------------------------------------------------------
   Note: _cons estimates baseline odds.

16 . *****used the code ib3.smoke_num to specify the category (never smoked) i want
   > ed as my reference category for the variable smoke_num

17 . ****permanently set category 3 as reference category for variable smoke_num

18 . fvset base 3 smoke_num

19 . *****interpretation: when all other variables are held constant, a one unit in
   > crease average glucose level, increases the odds of stroke by 0.5%. when all o
   > ther variables are held constant, the odds of stroke in individuals with hyper
   > tension is increased by 67.8% compared to those without hypertension. an addit
   > ional year of age increases the odds of stroke by 7.1%. average blood glucose,
   >  hypertension, and age have p-valus <0.05, thus these finndings are not a resu
   > lt of chance. They are statistically significant.

20 . ****adjusting for gender, bmi, average glucose level, presence of heart diseas
   > e, hypertension, and age, there was statistically significant relationship bet
   > weensmokers/former smokers and stroke when compared with individuals who had n
   > ever smoked.

21 . log close
         name:  <unnamed>
          log:  C:\Users\USER\Downloads\heart disease\multivariate logistics.smcl
     log type:  smcl
    closed on:  15 Jan 2026, 11:22:12
```