Assignment2
STAT229
Name: Zihao Huang
Date: 1/30/2017

1.

```
. tab sta inf
```

STA	INF 0	1	Total
0	100	60	160
1	16	24	40
Total	116	84	200

The log-odds is (100*24) / (60*16) = 2.5

```
. logistic sta inf

Logistic regression                             Number of obs     =        200
                                                LR chi2(1)        =       6.58
                                                Prob > chi2       =     0.0103
Log likelihood = -96.792669                     Pseudo R2         =     0.0329
```

sta	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf. Interval]	
inf	2.5	.9042723	2.53	0.011	1.230418	5.079573
_cons	.16	.0430813	-6.81	0.000	.0943902	.2712146

The log-odds is 2.5

```
. logit sta inf

Iteration 0:   log likelihood = -100.08048
Iteration 1:   log likelihood = -96.853843
Iteration 2:   log likelihood = -96.792686
Iteration 3:   log likelihood = -96.792669
Iteration 4:   log likelihood = -96.792669

Logistic regression                             Number of obs     =        200
                                                LR chi2(1)        =       6.58
                                                Prob > chi2       =     0.0103
Log likelihood = -96.792669                     Pseudo R2         =     0.0329
```

sta	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
inf	.9162908	.3617089	2.53	0.011	.2073543	1.625227
_cons	-1.832582	.2692582	-6.81	0.000	-2.360318	-1.304845

The slope coefficient of INF is e^0.9163=2.500 and Std is 0.3617

Also,
SD = [sqrt (1/100+1/16+1/24+1/60)] = 0.3617

And the 95% confident interval of OR of INF is (e^0.2074, e^1.6252)
1.23 =< OR =< 5.08

2.

LOW	RACE 1	2	3	Total
0	**73**	**15**	**42**	**130**
1	**23**	**11**	**25**	**59**
Total	**96**	**26**	**67**	**189**

For race=1 is reference group,
The OR for race=1 and race=2 is (73*11) / (15*23) = 2.3275 (e^0.8448)
The OR for race=1 and race=3 is (73*25) / (23*42) = 1.8892 (e^0.6362)

```
. logistic low i.race

Logistic regression                             Number of obs     =        189
                                                LR chi2(2)        =       5.01
                                                Prob > chi2       =     0.0817
Log likelihood = -114.83082                     Pseudo R2         =     0.0214
```

low	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
race						
2	2.327536	1.078613	1.82	0.068	.9385072	5.772385
3	1.889234	.6571342	1.83	0.067	.9554577	3.735597
_cons	.3150685	.0753382	-4.83	0.000	.1971825	.503433

```
. logit low i.race, level(90)

Iteration 0:   log likelihood =   -117.336
Iteration 1:   log likelihood = -114.84273
Iteration 2:   log likelihood = -114.83082
Iteration 3:   log likelihood = -114.83082

Logistic regression                             Number of obs     =        189
                                                LR chi2(2)        =       5.01
                                                Prob > chi2       =     0.0817
Log likelihood = -114.83082                     Pseudo R2         =     0.0214

------------------------------------------------------------------------------
         low |      Coef.   Std. Err.      z    P>|z|     [90% Conf. Interval]
-------------+----------------------------------------------------------------
        race |
          2  |   .8448103   .4634141     1.82   0.068     .0825619    1.607059
          3  |   .6361714    .347831     1.83   0.067     .0640403    1.208303
             |
       _cons |  -1.154965   .2391169    -4.83   0.000    -1.548278   -.7616529
------------------------------------------------------------------------------
```

The log-odds of low by race2 is e^0.8448 = 2.3275
The 95% confident interval of OR for race2 is (e^0.0826, e^1.6071), that is (1.0861, 4.9883)
The log-odds of low by race3 is e^0.6362 = 1.8893.
The 95% confident interval of OR for race3 is (e^0.0640, e^1.2083), that is (1.0661, 3.3478)

In 90% confidence (because the p-value is smaller than 0.10 but larger than 0.05),

```
Covariance matrix of coefficients of logistic model

             |        low
        e(V) |       race       _cons
-------------+------------------------
low          |
        race |  .02937799
       _cons | -.05714275   .13626626
```

3.

```
. logit sta crn

Iteration 0:   log likelihood = -100.08048
Iteration 1:   log likelihood = -97.599816
Iteration 2:   log likelihood = -97.368391
Iteration 3:   log likelihood = -97.368374
Iteration 4:   log likelihood = -97.368374

Logistic regression                               Number of obs     =        200
                                                  LR chi2(1)        =       5.42
                                                  Prob > chi2       =     0.0199
Log likelihood = -97.368374                       Pseudo R2         =     0.0271

------------------------------------------------------------------------------
         sta |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         crn |   1.219757   .5038556     2.42   0.015     .2322177    2.207296
       _cons |   -1.53821   .1948369    -7.89   0.000    -1.920084   -1.156337
------------------------------------------------------------------------------
```

```
. logit sta crn age

Iteration 0:   log likelihood = -100.08048
Iteration 1:   log likelihood = -94.672038
Iteration 2:   log likelihood =   -94.3031
Iteration 3:   log likelihood = -94.302294
Iteration 4:   log likelihood = -94.302294

Logistic regression                               Number of obs     =        200
                                                  LR chi2(2)        =      11.56
                                                  Prob > chi2       =     0.0031
Log likelihood = -94.302294                       Pseudo R2         =     0.0577

------------------------------------------------------------------------------
         sta |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         crn |   1.019856   .5149244     1.98   0.048     .0106231     2.02909
         age |   .0249915   .0107234     2.33   0.020      .003974     .046009
       _cons |  -3.029875   .7000292    -4.33   0.000    -4.401907   -1.657843
------------------------------------------------------------------------------
```

```
. logit sta age crn i.crn#c.age

Iteration 0:   log likelihood = -100.08048
Iteration 1:   log likelihood = -94.160869
Iteration 2:   log likelihood = -93.683315
Iteration 3:   log likelihood = -93.681076
Iteration 4:   log likelihood = -93.681076

Logistic regression                             Number of obs     =        200
                                                LR chi2(3)        =      12.80
                                                Prob > chi2       =     0.0051
Log likelihood = -93.681076                     Pseudo R2         =     0.0639
```

sta	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.029242	.011725	2.49	0.013	.0062613	.0522226
crn	3.573101	2.322261	1.54	0.124	-.978447	8.124649
crn#c.age						
1	-.0380925	.0340579	-1.12	0.263	-.1048447	.0286598
_cons	-3.297927	.7705345	-4.28	0.000	-4.808147	-1.787707

From these three models: STA explained by CRN, STA explained by CRN and AGE, STA explained by CRN and AGE and interaction of CRN and AGE, for model 1 and model 2, G=-2*(-93.681076-(-94.302294)), is larger than p-value for chi-square 0.05, means there is no difference between model1 and model 2. And the difference of CRN is (1.219757-1.019856)/1.019856 = 0.1960=19.6%, which means the change is nearly 20%, it needs adjustment. It can be seen that AGE make CRN's p-value larger which means it may be a confounder. And from the statistical interaction, we knew that the p-value of interaction between AGE and CRN is 0.263, which is large and it shows that in 95% confidence it is not significant in this model. So, there is no statistical interaction between these two variables.

4.

(a)

```
. logit death inh_inj

Iteration 0:   log likelihood = -422.70909
Iteration 1:   log likelihood =  -369.7569
Iteration 2:   log likelihood = -352.08759
Iteration 3:   log likelihood = -345.83117
Iteration 4:   log likelihood =  -345.8305
Iteration 5:   log likelihood =  -345.8305

Logistic regression                             Number of obs     =      1,000
                                                LR chi2(1)        =     153.76
                                                Prob > chi2       =     0.0000
Log likelihood =  -345.8305                     Pseudo R2         =     0.1819

------------------------------------------------------------------------------
       death |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     inh_inj |   2.692546   .2189964    12.29   0.000     2.263321    3.121771
       _cons |  -2.327903   .1186192   -19.63   0.000    -2.560392   -2.095414
------------------------------------------------------------------------------

. logit death inh_inj age

Iteration 0:   log likelihood = -422.70909
Iteration 1:   log likelihood = -301.37879
Iteration 2:   log likelihood = -269.42894
Iteration 3:   log likelihood =  -266.7758
Iteration 4:   log likelihood =    -266.77
Iteration 5:   log likelihood =    -266.77

Logistic regression                             Number of obs     =      1,000
                                                LR chi2(2)        =     311.88
                                                Prob > chi2       =     0.0000
Log likelihood =    -266.77                     Pseudo R2         =     0.3689

------------------------------------------------------------------------------
       death |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     inh_inj |   2.923704   .2686788    10.88   0.000     2.397104    3.450305
         age |   .0587824   .0055477    10.60   0.000     .0479091    .0696557
       _cons |  -5.002426    .338955   -14.76   0.000    -5.666765   -4.338086
------------------------------------------------------------------------------
```

From these two models: death explained by inh_inj ,and death explained by inh_inj and age, Both inh_inj and age in these two models are significant, the percentage of inh_inj from model 1 to model 2 is (2.6925-2.9237)/2.9237=-0.0791=7.91%, and for these two model, G=-2*(-266.77-(-345.8305)), which is much smaller than chi-square (1)'s p-value 0.001, thus, these two models are significant.