Part 1: Logistic Regression Analysis for cross-sectional data

Data: Myocardial Infraction.
Variables: Oral Contraceptive indicator (0: No, 1:Yes)
Age in Years
Smoking Indicator (0: No, 1: Yes)
MI (0: No, 1: Yes)
Question: Is MI related to OC use, age and smoking?
Method: Logistic regression

Exploratory Analysis

Relation between smoking and MI

```
. tabulate  smoking MI

           |          MI
   smoking |         0          1 |     Total
-----------+----------------------+----------
         0 |       117         17 |       134
         1 |        40         26 |        66
-----------+----------------------+----------
     Total |       157         43 |       200
```

```
. cci 26 17 40 117

                 |   Exposed   Unexposed  |     Total     Proportion
                 |                        |                  Exposed
-----------------+------------------------+----------------------
           Cases |        26          17  |        43      0.6047
        Controls |        40         117  |       157      0.2548
-----------------+------------------------+----------------------
           Total |        66         134  |       200      0.3300
                 |                        |
                 |      Point estimate    |  [95% Conf. Interval]
                 |------------------------+----------------------
      Odds ratio |         4.473529       |  2.215584    9.031184
 Attr. frac. ex. |         .7764629       |  .5486518    .8892725
 Attr. frac. pop |         .4694892       |
                 +-------------------------------------------------
                             chi2(1) =    18.69  Pr>chi2 = 0.0000
```

OR: 4.47, CI: (2.21, 9.03), p-value: 0.00 (Ho: OR = 1)

Conclusion:

Relation between Oral conceptive use and MI

```
. tabulate   OC MI

           |          MI
        OC |         0          1 |     Total
-----------+----------------------+----------
         0 |        93          9 |       102
         1 |        64         34 |        98
-----------+----------------------+----------
     Total |       157         43 |       200
```

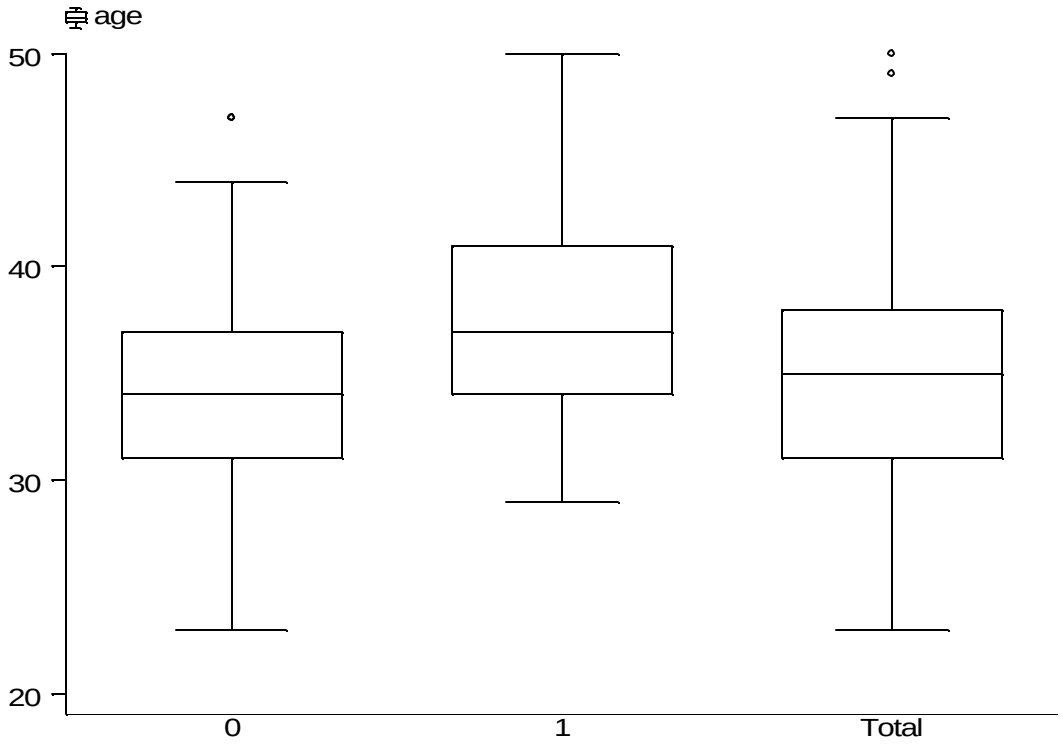
```
. cci 34 9 64 93

                                                        Proportion
                |   Exposed   Unexposed  |     Total     Exposed
----------------+------------------------+----------------------
          Cases |        34           9  |        43      0.7907
       Controls |        64          93  |       157      0.4076
----------------+------------------------+----------------------
          Total |        98         102  |       200      0.4900
                |                        |
                |      Point estimate    |  [95% Conf. Interval]
                |------------------------+----------------------
     Odds ratio |         5.489583       |  2.496017     12.0376
 Attr. frac. ex. |         .8178368       |  .5993616    .9169269
 Attr. frac. pop |         .6466617       |
                +-------------------------------------------------
                           chi2(1) =    19.82  Pr>chi2 = 0.0000
```

OR: 5.4 , CI: (2.30, 12.04), p-value: 0.00 (Ho: OR = 1)

Age and MI

```
. graph age, box by(MI) total s(o) ylab
```



```
. summ age if MI == 1

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+--------------------------------------------------------
     age |      43    37.90698   5.079435         29         50

. summ age if MI == 0

Variable |     Obs        Mean   Std. Dev.       Min        Max
---------+--------------------------------------------------------
     age |     157    34.12102   4.682688         23         47
```

Now let's perform a two sample t test.

```
. ttesti 43 37.907 5.079 157 34.121 4.683

------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |      43      37.907    .7745402       5.079    36.34391    39.47009
       y |     157      34.121    .3737441       4.683    33.38275    34.85925
---------+--------------------------------------------------------------------
combined |     200    34.93499    .3540305    5.006747    34.23686    35.63312
---------+--------------------------------------------------------------------
    diff |               3.786    .8209683                2.167036    5.404964
------------------------------------------------------------------------------
Degrees of freedom: 198

                     Ho: mean(x) - mean(y) = diff = 0

     Ha: diff < 0                 Ha: diff ~= 0                 Ha: diff > 0
       t =   4.6116                 t =   4.6116                 t =   4.6116
   P < t =   1.0000           P > |t| =   0.0000           P > t =   0.0000
```

Now lets perform logistic regresion analysis.

```
. glm MI age  smoking OC, f(bin) l(logit)

Iteration 1 : deviance =  156.9761
Iteration 2 : deviance =  150.7970
Iteration 3 : deviance =  150.3779
Iteration 4 : deviance =  150.3748
Iteration 5 : deviance =  150.3748

Residual df  =       196                                 No. of obs =       200
Pearson X2   =  177.5298                                 Deviance   =  150.3748
Dispersion   =  .9057645                                 Dispersion =  .7672184

Bernoulli distribution, logit link
------------------------------------------------------------------------------
      MI |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1625556   .0445356      3.650   0.000       .0752673    .2498438
 smoking |   1.812237    .429383      4.221   0.000       .9706619    2.653812
      OC |   1.979896   .4696779      4.215   0.000       1.059344    2.900448
   _cons |  -9.114047   1.757031     -5.187   0.000      -12.55777   -5.670329
------------------------------------------------------------------------------
```

```
. glm, eform

Residual df  =       196                               No. of obs =       200
Pearson X2   =  177.5298                               Deviance   =  150.3748
Dispersion   =  .9057645                               Dispersion =  .7672184

Bernoulli distribution, logit link
------------------------------------------------------------------------------
      MI | Odds Ratio   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   1.176514   .0523968      3.650   0.000       1.078172    1.283825
 smoking |   6.124132   2.629598      4.221   0.000       2.639691     14.2081
      OC |   7.241988   3.401402      4.215   0.000       2.884478    18.18228
------------------------------------------------------------------------------
```

Part 2: Logistic Regression Analysis for longitudional data with random effects.

Data: Wheezing

Model:

$$\text{logit } Pr(Y_{ij} = 1| U_i) = \beta_0 + U_i + b\,\mathbf{X}$$

We assume that conditional on the unobservable responses U_i, we have independent responses from a distribution in exponential family.

Interpretation of parameters: Consider the followig model, logit for the ith individual in the study.

$$\text{logit } Pr(Y_i = 1| U_i) = \beta^*_0 + U_i + \beta^*_1 x_{ij}$$

where x_{ij} is 1 if child I is vitamin A deficient and 0 otherwise. Distribution of U_i is normal with mean 0 and unknown variance ν^2. In this example β^*_0 is the log odds of respiratory infection for a typical child with random effect $U_i = 0$. The parameter β_1^* is the log odds for respiratory infection when a child is deficient relative to when that same child is not. The variance ν^2 represents the degree of heterogeneity across children in the propensity of disease not attributible to x.

Let's fit the ramdom effects model with one covariate at time

Age and weezing

```
.  xtlogit  wee  age, i(Id) re

Fitting comparison model:

Iteration 0:   log likelihood = -60.215168
Iteration 1:   log likelihood = -59.821289
Iteration 2:   log likelihood = -59.820619

Fitting full model:

rho =  0.0     log likelihood = -59.820619
```

```
rho =  0.1     log likelihood = -59.534747
rho =  0.2     log likelihood = -59.313489
rho =  0.3     log likelihood = -59.174587
rho =  0.4     log likelihood = -59.144047
rho =  0.5     log likelihood = -59.264113
Iteration 0:   log likelihood = -59.144047
Iteration 1:   log likelihood = -59.053072
Iteration 2:   log likelihood = -59.052345
Iteration 3:   log likelihood = -59.052344

Random-effects logit                            Number of obs      =       100
Group variable (i) : Id                         Number of groups   =        32

Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       3.1
                                                               max =         4

                                                Wald chi2(1)       =      0.92
Log likelihood  = -59.052344                    Prob > chi2        =    0.3385

------------------------------------------------------------------------------
     wee |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |  -.2058959   .2151199     -0.957   0.339      -.6275231    .2157313
   _cons |    1.13757    2.24621      0.506   0.613      -3.264921    5.540061
---------+--------------------------------------------------------------------
/lnsig2u |  -.3512171   1.122664     -0.313   0.754      -2.551598    1.849164
---------+--------------------------------------------------------------------
 sigma_u |   .8389463   .4709274                          .2792078    2.520814
     rho |   .4130873   .2721856                          .0723192    .8640289
------------------------------------------------------------------------------
Likelihood ratio test of rho=0:      chi2(1) =     1.54   Prob > chi2 = 0.2151
```

Now all the covariates

```
. xtlogit  wee smk2  age City2, i(Id) re

Fitting comparison model:

Iteration 0:   log likelihood = -60.215168
Iteration 1:   log likelihood = -58.927241
Iteration 2:   log likelihood = -58.919236
Iteration 3:   log likelihood = -58.919235

Fitting full model:

rho =  0.0     log likelihood = -58.919235
rho =  0.1     log likelihood =  -58.62208
rho =  0.2     log likelihood = -58.391529
rho =  0.3     log likelihood = -58.244856
rho =  0.4     log likelihood = -58.208104
rho =  0.5     log likelihood = -58.324007
Iteration 0:   log likelihood = -58.208104
Iteration 1:   log likelihood = -58.098593
Iteration 2:   log likelihood = -58.097518
Iteration 3:   log likelihood = -58.097517

Random-effects logit                            Number of obs      =       100
Group variable (i) : Id                         Number of groups   =        32
```

```
Random effects u_i ~ Gaussian                   Obs per group: min =         1
                                                               avg =       3.1
                                                               max =         4

                                                Wald chi2(3)       =      2.62
Log likelihood  = -58.097517                    Prob > chi2        =    0.4548

------------------------------------------------------------------------------
         wee |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
        smk2 |  -.6623059   .5200883     -1.273   0.203       -1.68166    .3570485
         age |   -.198715   .2181834     -0.911   0.362      -.6263465    .2289166
       City2 |   .2836571   .5794732      0.490   0.624      -.8520895    1.419404
       _cons |   1.216874   2.287129      0.532   0.595      -3.265817    5.699565
-------------+----------------------------------------------------------------
    /lnsig2u |  -.3038675    1.09995     -0.276   0.782       -2.45973    1.851995
-------------+----------------------------------------------------------------
     sigma_u |   .8590452   .4724534                           .292332    2.524385
         rho |   .4246123   .2687362                          .0787299    .8643612
------------------------------------------------------------------------------
Likelihood ratio test of rho=0:      chi2(1) =     1.64   Prob > chi2 = 0.1999
```

What are the interpretations of the parameters?