Chapter 11

Analysis of Correlated Data

Aim: Upon completing the chapter, the learner should be able to fit a linear regression with correlated data.

11.1 Regression Models with Correlated Data

When the measurements are part of a group of individuals, then it is possible that these measurements will be correlated. For example, patients receiving medical care in a hospital (private or public) and who are assigned to a doctor (generalist or specialist), it is possible that the treatments that these patients receive will be very similar. The patients are observed nested within certain type of doctors and hospitals.

Another example of correlated data is in cross-sectional studies, when a complex sampling design of households is used (Figure 11.1); the possibility exists that the responses of the people who reside on the same census block about lifestyle habits will be similar.

Another situation in which there might be correlated data is when repeated measurements are made; for example, the different weights of a single child at different visits to his or her pediatrician. There are various alternatives for analyzing these data, and which alternatives to use will depend on the objectives of the investigator. One alternative is to analyze the independent measurements separately, which results in a loss of information and statistical power. For example, assume the following data with regard to 20 different children, each one visiting his or her pediatrician three different times, being weighed at each visit, and information on whether they practice a sport on a regular basis (1 = yes vs. 0 = no):

	id	weight1	weight2	weight3	sport
1.	1	66	67	68	0
2.	2	71	71	65	0
3.	3	70	66	62	0
4.	4	64	62	66	1
5.	5	67	66	68	1
6.	6	65	64	65	1
7.	7	67	67	63	0
8.	8	65	66	66	1
9.	9	69	70	68	0
10.	10	63	62	63	1
11.	11	61	60	60	1
12.	12	66	68	68	1
13.	13	68	68	70	0
14.	14	67	69	65	0
15.	15	65	67	63	1
16.	16	64	62	64	1
17.	17	65	64	55	0
18.	18	65	65	66	0
19.	19	64	63	62	1
20.	20	67	66	63	0

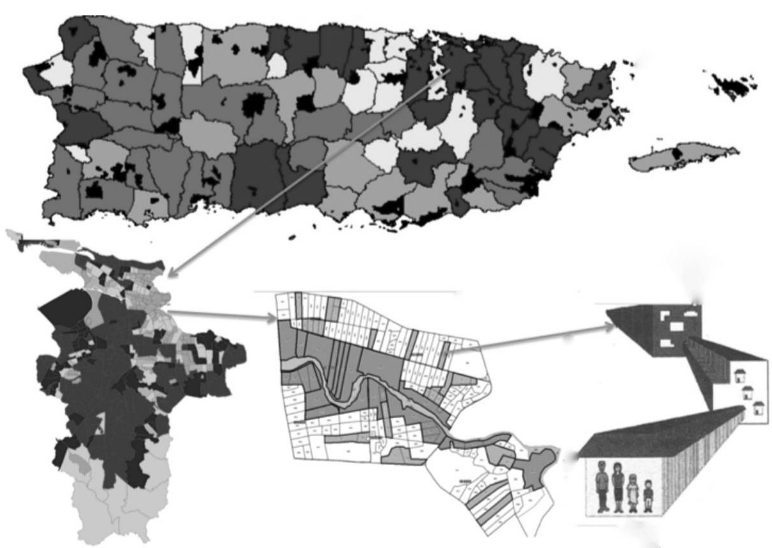


Figure 11.1 Sampling design in Puerto Rico.

To perform the analysis with independent measurements, assuming that the objective is to compare the average weight by type of sport, we could carry out a simple linear regression analysis of the weight of each child, using the following command lines at each visit:

For the first visit:

```
reg weight1 sport
```

Output

```
    Source |      SS     df       MS       Number of obs   =      20
-----------+------------------------       F(1, 18)        =   14.20
     Model |    48.05     1        48.05   Prob > F        = 0.0014
  Residual |     60.9    18   3.38333333   R-squared       = 0.4410
-----------+------------------------       Adj R-squared   = 0.4100
     Total |   108.95    19   5.73421053   Root MSE        = 1.8394

--------------------------------------------------------------------
   weight1 | Coef. Std. Err.     t  P>|t|   [95% Conf. Interval]
-----------+--------------------------------------------------------
     sport | -3.1   .8225975   -3.77 0.001   -4.828213   -1.371787
     _cons | 67.5   .5816643  116.05 0.000    66.27797    68.72203
--------------------------------------------------------------------
```

The results show a significant effect of the predictor sport on mean weight at visit 1 (P-value = .001). The estimated regression coefficient for the predictor *sport* is the difference between the mean weights by sport in visit 1. The following command line can be used to compute the observed mean weight by sport:

```
table sport, c(mean weight1)
```

Output

```
------------------------
   sport | mean(weight1)
---------+--------------
       0 |          67.5
       1 |          64.4
------------------------
```

The difference in the mean weights at visit 1 is −3.1, so the children who practice regularly a sport weigh less, on average, than those who do not practice regularly a sport. To explore the differences in mean weight in each visit, a line can be drawn between the estimated weights from a linear regression model by sport. For example, in the first visit the following Stata commands for visualizing this line can be used to create Figure 11.2:

```
predict weight1exp
twoway (line weight1exp sport, sort), ytitle(Mean weight)
xtitle(Sport) xlabel(0(1)1) legend(off)
```

We repeat the previous steps for the second visit:

```
reg weight2 sport
```

Output

```
    Source |     SS      df      MS      Number of obs  =       20
-----------+-----------------------      F(1, 18)       =     9.24
     Model |    54.45     1       54.45  Prob > F       =   0.0071
  Residual |    106.1    18 5.89444444   R-squared      =   0.3391
-----------+-----------------------      Adj R-squared  =   0.3024
     Total |   160.55    19        8.45  Root MSE       =   2.4278

------------------------------------------------------------------------
  weight2   |Coef.  Std. Err.   t     P>|t|   [95% Conf. Interval]
------------+-----------------------------------------------------------
    sport   |-3.3  1.085766  -3.04  0.007  -5.581111  -1.018889
    _cons   |67.3  .7677529  87.66  0.000   65.68701   68.91299
------------------------------------------------------------------------
```

```
table sport, c(mean weight2)
```

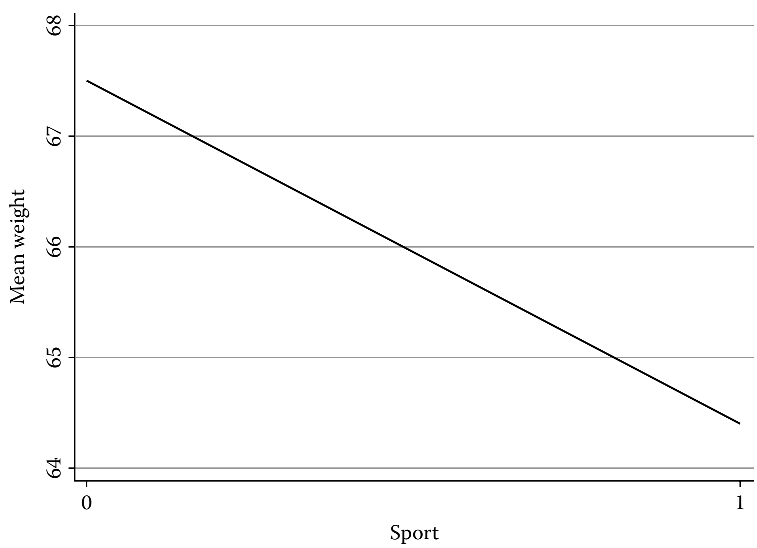


Figure 11.2 Mean weight in the first visit.

Output

```
-------------------------
    sport |  mean(weight2)
----------+--------------
        0 |          67.3
        1 |            64
-------------------------
```

The results also show a significant effect of the predictor sport on mean weight at visit 2 (P-value = .007). The difference in the mean weights at the second visit is −3.3; children who practice regularly a sport weight less, on average, than those who do not practice regularly a sport. To draw the estimated weight by sport at visit 2, the following Stata commands are used to create Figure 11.3:

```
predict weight2exp
twoway (line weight2exp sport, sort), ytitle(Mean weight)
xtitle(Sport) xlabel(0(1)1) legend(off)
```

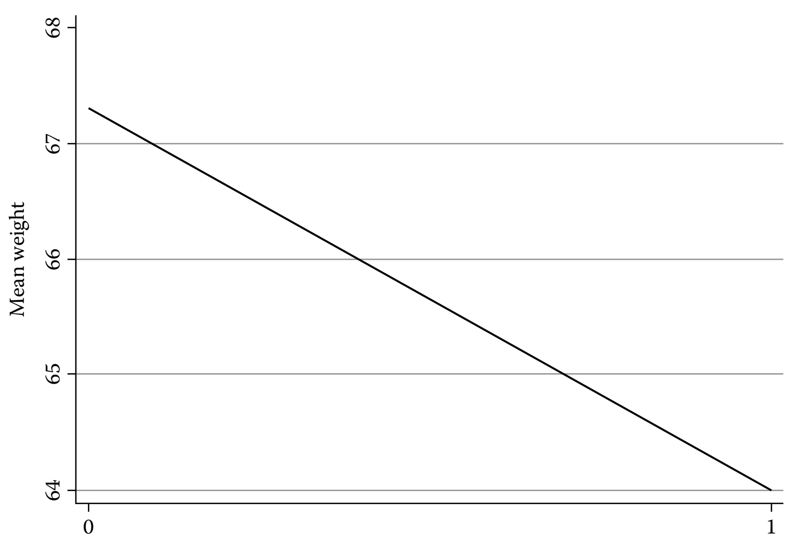


Figure 11.3 Mean weight in the second visit.

The Stata commands on the third visit are:

```
reg weight3 sport
```

Output

```
      Source |       SS       df       MS              Number of obs =      20
-------------+------------------------------           F(1, 18)      =    0.00
       Model |        0     1            0             Prob > F      =  1.0000
    Residual |      219    18   12.1666667             R-squared     =  0.0000
-------------+------------------------------           Adj R-squared = -0.0556
       Total |      219    19   11.5263158             Root MSE      =  3.4881

------------------------------------------------------------------------------
     weight3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       sport |          0   1.559915     0.00   1.000    -3.277259    3.277259
       _cons |       64.5   1.103026    58.48   0.000     62.18263    66.81737
------------------------------------------------------------------------------
```

```
table sport, c(mean weight3)
```

Output

```
-------------------------
    sport | mean(weight3)
----------+--------------
        0 |          64.5
        1 |          64.5
-------------------------
```

The results do not show a significant effect of the predictor sport on mean weight at visit 3 (P-value > .1). There is no difference in the mean weights at the third visit. To draw the estimated weight by sport on the third visit, the following Stata commands are used to create Figure 11.4:

```
predict weight3exp
twoway (line weight3exp sex, sort), ytitle(Mean weight)
xtitle(Sport) xlabel(0(1)1) legend(off)
```

11.2 Mixed Models

The other alternative for analyzing correlated data is to use mixed models or multilevel models that allow us to correct the statistical relationship due to the potential correlation between measurements. For example, the simplest scheme of correlation is the repeated measures study; the same subject is measured several times, as illustrated in Figure 11.5.

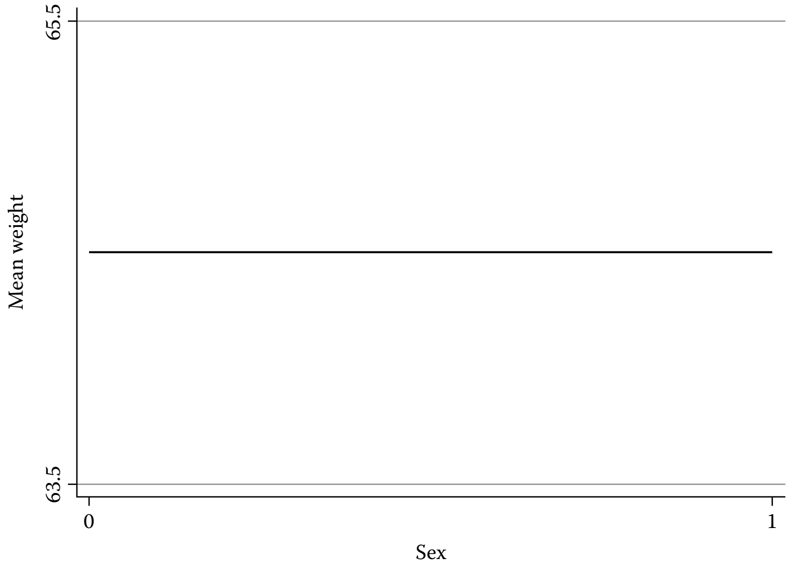
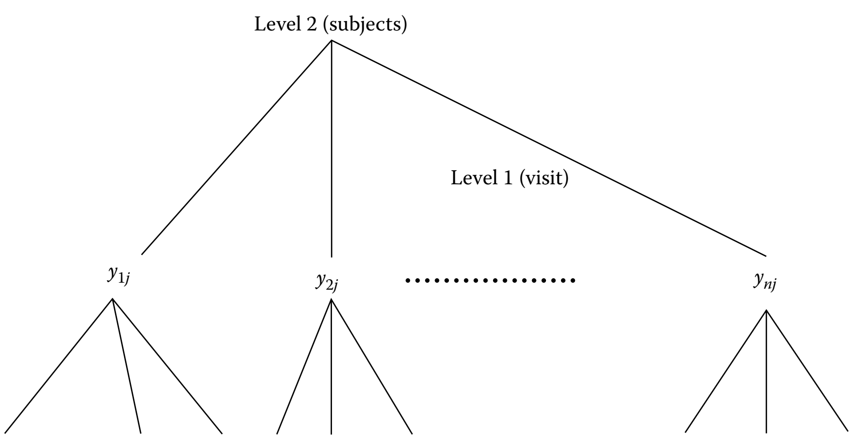


Figure 11.4 Mean weight in the third visit.



y_{ij} indicates the value of Y in the jth visit for the ith subject.

Level 1 refers to the set of weights of each subject on different visits.

Level 2 refers to the set of subjects.

Figure 11.5 Repeated measures of weight in the same subject.

Mixed models can be expressed in various forms to explain the expected value of the *main outcome* (Y) of the study. The construction of these models depends on the following variations:

1. $\bar{Y}_i$ may be different between subjects at baseline (variability between measurements).
2. $\bar{Y}_i$ may change according to the conditions in each subject (variability within measurements).

When an association is established between a continuous random variable, Y, and a quantitative variable, X, through a simple linear regression model, $\mu_{Y|X} = \beta_0 + \beta_1 * X$, in a mixed model approach, the intercept (β_0) and the slope (β_1) could be fixed or random. The following table indicates the possible combinations of these alternatives:

Intercept (β_0)	*Slope* (β_1)
Fixed	Fixed
Fixed	Random
Random	Fixed
Random	Random

Based on the previous example of the estimated weights by sport, we need to identify the possible patterns of the linear relationships. According to the previous graphs, the respective patterns in the lines suggest a model with a random intercept considering the first two visits, similar slopes but different intercept; it means that the difference in the mean weight, between those who practice regularly a sport and those who do not, is independent of the visit. However, if the three visits are considered, a model with random intercept and slope is suggested; the difference in the mean weight, between those who practice sport and those who do not, is not independent of the visit.

11.3 Random Intercept

The definition of a linear mixed model with a random intercept is as follows (Rabe-Hesketh and Skrondal, 2005):

$$Y_{ij} = \beta_{0i} + \beta_1 * X_{ij} + e_{ij}$$

where:

$\beta_{0i} = \gamma_{00} + U_{0i}$ (random coefficient)

β_1 indicates the coefficients (fixed) associated to the predictor X

$U_{0i} \sim N\left(0, \sigma_{U_0}^2\right)$

γ_{00} indicates the average intercept in a design with two levels

σ_{u0}^2 indicates the variance of the intercept between subjects

The variance of Y_{ij} conditional on the value of X_{ij} is given by the following expression:

$$\text{var}\left(Y_{ij}|X_{ij}\right) = \text{var}\left(U_{0j}\right) + \text{var}\left(\varepsilon_{ij}\right) = \sigma_{u0}^2 + \sigma_\varepsilon^2$$

where σ_ε^2 indicates the variance of Y within the subjects (variance of the residuals).

Using the data of the previous example, the correlation between two different visits ($j \neq j'$) of the ith subject is calculated through the covariance expression, as follows:

$$\text{cov}\left(Y_{ij}, Y_{ij'}|X_{ij}, X_{ij'}\right) = \text{var}\left(U_{0j}\right) = \sigma_{u0}^2$$

Therefore, the correlation between the measurements of Y in a subject, when Y is fitted by X, is defined with the following equation:

$$\rho\left(Y_{ij} \mid X_{ij}\right) = \frac{\text{cov}\left(Y_{ij}, Y_{ij'} \mid X_{ij}, X_{ij'}\right)}{\text{var}\left(Y_{ij} \mid X_{ij}\right)} = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_\varepsilon^2}$$

This measurement is known as an intraclass correlation coefficient, which does not change according to the value of X.

11.4 Using the *mixed* and *gllamm* Commands with a Random Intercept

To run the mixed model with the assumption of a random intercept, we need the database to be in the long format. The following command needed to get a long database from a wide database is:

```
reshape long weight, i(id) j(visit)
```

After submitting this command, the database is changed as follows:

```
     +-----------------------------+
     | id   visit   weight   sport |
     |-----------------------------|
  1. |  1       1       66       0 |
  2. |  1       2       67       0 |
  3. |  1       3       68       0 |
  4. |  2       1       71       0 |
  5. |  2       2       71       0 |
     |-----------------------------|
  6. |  2       3       65       0 |
  7. |  3       1       70       0 |
  8. |  3       2       66       0 |
  9. |  3       3       62       0 |
 10. |  4       1       64       1 |
     |-----------------------------|
 11. |  4       2       62       1 |
 12. |  4       3       66       1 |
 13. |  5       1       67       1 |
 14. |  5       2       66       1 |
 15. |  5       3       68       1 |
     |-----------------------------|
 16. |  6       1       65       1 |
 17. |  6       2       64       1 |
 18. |  6       3       65       1 |
 19. |  7       1       67       0 |
 20. |  7       2       67       0 |
     |-----------------------------|
 21. |  7       3       63       0 |
 22. |  8       1       65       1 |
 23. |  8       2       66       1 |
 24. |  8       3       66       1 |
 25. |  9       1       69       0 |
     |-----------------------------|
 26. |  9       2       70       0 |
 27. |  9       3       68       0 |
 28. | 10       1       63       1 |
 29. | 10       2       62       1 |
 30. | 10       3       63       1 |
     |-----------------------------|
 31. | 11       1       61       1 |
 32. | 11       2       60       1 |
 33. | 11       3       60       1 |
 34. | 12       1       66       1 |
 35. | 12       2       68       1 |
     |-----------------------------|
```

```
36. | 12        3        68        1 |
37. | 13        1        68        0 |
38. | 13        2        68        0 |
39. | 13        3        70        0 |
40. | 14        1        67        0 |
    |------------------------------|
41. | 14        2        69        0 |
42. | 14        3        65        0 |
43. | 15        1        65        1 |
44. | 15        2        67        1 |
45. | 15        3        63        1 |
    |------------------------------|
46. | 16        1        64        1 |
47. | 16        2        62        1 |
48. | 16        3        64        1 |
49. | 17        1        65        0 |
50. | 17        2        64        0 |
    |------------------------------|
51. | 17        3        55        0 |
52. | 18        1        65        0 |
53. | 18        2        65        0 |
54. | 18        3        66        0 |
55. | 19        1        64        1 |
    |------------------------------|
56. | 19        2        63        1 |
57. | 19        3        62        1 |
58. | 20        1        67        0 |
59. | 20        2        66        0 |
60. | 20        3        63        0 |
    +------------------------------+
```

Subsequently, the ***mixed*** command is used for the first two visits as can be seen below:

```
mixed weight sport if visit < 3, || id:,  stddev
```

Output

```
Mixed-effects ML regression          Number of obs      =        40
Group variable: id                   Number of groups   =        20

                                     Obs per group:
                                                   min =         2
                                                   avg =       2.0
                                                   max =         2
                                     Wald chi2(1)       =     14.12
Log likelihood = -77.965175          Prob > chi2        =    0.0002
-------------------------------------------------------------------------------
```

```
    weight | Coef. Std. Err.      z  P>|z|  [95% Conf. Interval]
-------------+----------------------------------------------------
     sport | -3.2   .8514694   -3.76 0.000 -4.868849  -1.531151
     _cons | 67.4   .6020798  111.95 0.000  66.21995   68.58005
-------------------------------------------------------------------

-------------------------------------------------------------------
Random-effects Parameters|Estimate Std. Err. [95% Conf. Interval]
-------------------------+-----------------------------------------
id: Identity             |
               sd(_cons) |1.746425 .3322952  1.202792  2.535768
-------------------------+-----------------------------------------
            sd(Residual) | 1.07238 .1695582  .7866148   1.46196
-------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 14.99 Prob >= chibar2 = 0.0001
```

The results indicate that there is a significant change in the expected weight by sport (P-value <.001), even after controlling for the effect between subjects ($\hat{\beta}_1 = -3.2$ 95% CI: −4.87, −1.53). The intraclass correlation coefficient is determined with the following estimates:

$$\hat{\sigma}_{U0} = 1.74$$

$$\hat{\sigma}_{\varepsilon} = 1.07$$

$$\hat{\rho} = \frac{1.74^2}{1.74^2 + 1.07^2} = .73$$

We can see that the average intercept is $\hat{\gamma}_{00} = 67.4$ and varies per visit ± 1.74.

Another option when running the mixed model with a random intercept is to use the ***gllamm*** command, which may be downloaded from www.gllamm.org. The result is the following, assuming that the data are in the long format:

```
gllamm weight sport if visit < 3, i(id) nip(20)
```

Output

```
gllamm model

log likelihood = -77.94849
```

```
------------------------------------------------------------------------------
 weight |   Coef.  Std. Err.    z    P>|z|  [95% Conf.  Interval]
---------+--------------------------------------------------------------------
  sport |-3.194099 .8346379   -3.83 0.000   -4.829959  -1.558239
  _cons | 67.38501 .570114   118.20 0.000    66.26761   68.50241
------------------------------------------------------------------------------

Variance at level 1
------------------------------------------------------------------------------
 1.1368944 (.37024967)

Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (id)

   var(1): 3.1128248 (1.2621798)
------------------------------------------------------------------------------
```

The results of both ***mixed*** and ***gllamm*** are similar, with a slight difference in the log-likelihood estimate and in the variance of random effects.

11.5 Using the *mixed* Command with Random Intercept and Slope

Another way to express a mixed model that explains the expected values of $Y\,(\mu_{ij})$ is to assume that the intercept and the slope are random variables, as follows:

$$\mu_{ij} = \beta_{0i} + \beta_{1i} * X_{ij}$$

where:

$\beta_{0i} = \gamma_{00} + U_{0i}$ (random coefficient)

β_{1i} indicates a random coefficient associated with the predictor X, which is defined as follows: $\gamma_{01} + U_{1i}$

$U_{0i} \sim N\left(0, \sigma^2_{U_0}\right)$

$U_{1i} \sim N\left(0, \sigma^2_{U_1}\right)$

γ_{00} indicates the average intercept in a design with two levels

γ_{01} indicates the average slope in a design with two levels

σ^2_{u0} indicates the variance of the intercept between subjects

σ^2_{u1} indicates the variance of the slope between subjects

In this mixed model, the variance of Y_{ij}, given X_{ij}, is enumerated by the following expression:

$$\text{var}\left(Y_{ij} \mid X_{ij}\right) = \sigma_{u0}^2 + 2\sigma_{01}X_{ij} + \sigma_{u1}^2 X_{ij}^2 + \sigma_\varepsilon^2,$$

where σ_{01} is the covariance between the intercept and the slope.

In addition, there is a possibility that there is a correlation between each of two different visits ($j \neq j'$) by the same subject, which is calculated from the covariance, as follows:

$$\text{Cov}\left(Y_{ij}, Y_{ij'} \mid X_{ij}, X_{ij'}\right) = \sigma_{u0}^2 + \sigma_{01}\left(X_{ij} + X_{ij'}\right) + \sigma_{u1}^2 * X_{ij} * X_{ij'}$$

As a consequence, the intraclass correlation coefficients with random intercept and random slope will be defined as follows:

$$\rho\left(Y_{ij} | X_{ij}\right) = \frac{\text{Cov}\left(Y_{ij}, Y_{ij'} | X_{ij}, X_{ij'}\right)}{\text{Var}\left(Y_{ij} | X_{ij}\right)} = \frac{\sigma_{u0}^2 + \sigma_{01}(X_{ij} + X_{ij'}) + \sigma_{u1}^2 X_{ij} * X_{ij'}}{\sigma_{u0}^2 + 2\sigma_{01}X_{ij} + \sigma_{u1}^2 X_{ij}^2 + \sigma_\varepsilon^2}$$

Therefore, this intraclass correlation coefficient will depend on the value of X.

When the intercept and slope are random in a mixed model, we can use the Stata command *mixed* with the previous database considering the three visits as follows:

```
mixed weight sport, || id: || sport:,  stddev
```

Output

```
Mixed-effects ML regression          Number of obs      =     60
---------------------------------------------------
                |  No. of   Observations per Group
Group Variable  |  Groups  Minimum  Average  Maximum
----------------+----------------------------------
             id |      20        3      3.0        3
          sport |      20        3      3.0        3
---------------------------------------------------

                                     Wald chi2(1)       =   5.16
Log likelihood = -140.95844          Prob > chi2        = 0.0231

------------------------------------------------------------------
  weight |  Coef.   Std. Err.     z   P>|z|  [95% Conf.  Interval]
---------+--------------------------------------------------------
   sport |-2.133333 .9387534  -2.27 0.023 -3.973256   -.2934105
   _cons | 66.43333 .6637989 100.08 0.000  65.13231    67.73436
------------------------------------------------------------------
```

```
-----------------------------------------------------------------------------
Random-effects Parameters|Estimate Std. Err. [95% Conf. Interval]
---------------------------+-------------------------------------------------
id: Identity               |
              sd(_cons)    |1.206639 18.1424  1.92e-13   7.58e+12
---------------------------+-------------------------------------------------
sport: Identity            |
              sd(_cons)    |1.206645 18.14232 1.92e-13   7.58e+12
---------------------------+-------------------------------------------------
           sd(Residual)    |2.1173   .2367064 1.700675   2.635989
-----------------------------------------------------------------------------
LR test vs. linear model: chi2(2) = 8.40 Prob > chi2 =  0.0150
```

The results indicate that there is a significant change in the average expected weight by sport (*P*-value = .023), even after controlling for the effect between subjects ($\hat{\gamma}_{01} = -2.13$, 95% CI: −3.97, −0.29). However, the average of the slopes associated with the sport predictor will vary ±1.21, so the pattern of change in the average weight will depend on each visit.

11.6 Mixed Models in a Sampling Design

Let us assume that we have the following information from a random sample of 88 subjects, which will be used to assess the association between hepatitis C virus infection (hcv: 1 = present; 0 = absent) and cocaine metabolite assay result (co: 1 = positive; 0 = negative), controlling for the variables age (age2: 1 = <45 years; 2 = ≥45 years) and residential block (block), selected randomly:

```
     +-------------------------+
     | block   co   hcv   age2 |
     |-------------------------|
  1. |     3    0     0      2 |
  2. |     4    1     1      1 |
  3. |     4    1     0      2 |
  4. |     1    0     0      2 |
  5. |     1    0     0      1 |
     |-------------------------|
  6. |     1    1     1      1 |
  7. |     4    1     0      1 |
  8. |     4    0     0      1 |
  9. |     2    0     0      1 |
 10. |     2    0     0      1 |
     |-------------------------|
 11. |     3    0     1      2 |
 12. |     2    0     0      1 |
 13. |     2    1     0      2 |
 14. |     4    0     0      1 |
 15. |     3    0     0      1 |
     |-------------------------|
```

```
16. |      1     1       0        2 |
17. |      1     0       0        1 |
18. |      4     1       0        1 |
19. |      2     1       1        1 |
20. |      2     1       0        1 |
    |-------------------------------|
21. |      2     0       0        1 |
22. |      2     1       0        1 |
23. |      3     0       0        1 |
24. |      3     1       0        1 |
25. |      3     0       0        1 |
    |-------------------------------|
26. |      3     0       0        2 |
27. |      1     1       0        1 |
28. |      1     1       0        1 |
29. |      4     0       0        2 |
30. |      4     0       0        1 |
    |-------------------------------|
31. |      4     1       1        2 |
32. |      4     1       1        1 |
33. |      1     0       0        1 |
34. |      2     0       1        1 |
35. |      2     0       0        1 |
    |-------------------------------|
36. |      3     0       0        2 |
37. |      3     0       1        1 |
38. |      2     1       0        2 |
39. |      4     0       0        1 |
40. |      4     0       1        1 |
    |-------------------------------|
41. |      1     0       0        1 |
42. |      4     0       0        2 |
43. |      4     0       0        1 |
44. |      1     0       0        1 |
45. |      2     1       0        2 |
    |-------------------------------|
46. |      2     0       0        1 |
47. |      1     1       1        1 |
48. |      1     1       0        1 |
49. |      3     0       0        1 |
50. |      3     0       0        2 |
    |-------------------------------|
51. |      2     0       1        1 |
52. |      2     1       0        1 |
53. |      2     0       0        1 |
54. |      4     0       0        1 |
55. |      4     0       0        1 |
    |-------------------------------|
```

```
56. |      4     1      1       1 |
57. |      4     0      0       1 |
58. |      4     0      0       1 |
59. |      3     1      0       1 |
60. |      3     1      0       2 |
    |-------------------------|
61. |      3     1      0       1 |
62. |      4     1      1       2 |
63. |      2     0      0       1 |
64. |      2     0      0       1 |
65. |      2     0      0       2 |
    |-------------------------|
66. |      1     0      0       2 |
67. |      2     0      0       1 |
68. |      2     0      0       2 |
69. |      4     0      0       1 |
70. |      2     0      0       2 |
    |-------------------------|
71. |      2     1      0       1 |
72. |      2     1      0       2 |
73. |      2     0      0       2 |
74. |      2     0      0       1 |
75. |      2     0      0       1 |
    |-------------------------|
76. |      3     1      0       2 |
77. |      3     0      0       2 |
78. |      3     0      0       1 |
79. |      1     0      0       1 |
80. |      1     0      0       2 |
    |-------------------------|
81. |      1     0      0       2 |
82. |      3     1      0       2 |
83. |      1     1      0       2 |
84. |      1     0      0       2 |
85. |      2     1      0       1 |
    |-------------------------|
86. |      2     0      0       1 |
87. |      2     1      0       2 |
88. |      1     1      0       1 |
    +-------------------------+
```

If the data are analyzed under the assumption that there is a possible correlation between the subjects that reside in the same block (random intercept), the syntax of the *gllamm* command line to estimate the prevalence ratio using a logistic regression model with the option *link(log)*, which is called log-binomial regression model, is as follows:

```
xi:gllamm hcv i.co i.age2,fam(bin) i(block) eform link(log)
```

Output

```
gllamm model

log likelihood = -34.37179

------------------------------------------------------------------------------
         hcv | exp(b)   Std. Err.   z    P>|z|   [95% Conf. Interval]
------------+-----------------------------------------------------------------
      _Ico_1 |2.834653 1.482238 1.99  0.046   1.017202   7.899373
    _Iage2_2 |.5385232 .3316394 -1.01 0.315   .1610675   1.800532
       _cons |.1036169 .0489457 -4.80 0.000   .0410533   .2615254
------------------------------------------------------------------------------
Variances and covariances of random effects
------------------------------------------------------------------------------

***level 2 (block)

   var(1): .0386266 (.20576199)
------------------------------------------------------------------------------
```

The results show that the prevalence of HCV infection among cocaine users is 2.83 (95% CI: 1.02, 7.90) times the prevalence of HCV infection among cocaine nonusers, adjusting for age and block of residence. This excess was statistically significant (*P*-value = .046).

There are other applications of multilevel modeling in health sciences that can be explored in Stata, including ordinal outcomes, count outcomes, and censored outcomes. These topics are beyond the scope of this book, but an extensive review of multilevel modeling can be found in Snijders and Bosker (2003), Leyland and Goldstein (2001), Twisk (2003), and Rabe-Hesketh and Skrondal (2005).