*STATA.OUTPUT -- Chapter 14

. clear
. *first example of a kappa statistic

. input r1 r2 freq

```
    r1 r2 freq
  1.  1 1 12
  2.  2 1  4
  3.  1 2  8
  4.  2 2 16
  5. end
```

. table r1 r2 [weight = freq]

```
----------------------
          |     r2
       r1 |    1     2
----------+-----------
        1 |   12     8
        2 |    4    16
----------------------
```

. kap r1 r2 [fw=freq]

```
             Expected
Agreement   Agreement     Kappa   Std. Err.         Z      Prob>Z
-----------------------------------------------------------------
  70.00%      50.00%     0.4000     0.1549       2.58      0.0049
```

. clear
. *second example of a kappa statistic

. input r1 r2 freq

```
    r1 r2 freq
  1.  1 1  3
  2.  2 1  7
  3.  3 1  4
  4.  1 2  2
  5.  2 2 14
  6.  3 2 19
  7.  2 3  1
  8.  3 3 42
  9.  4 3  1
 10.  3 4  2
 11.  4 4  3
 12. end
```

. table r1 r2 [weight = freq]

r1	r2: 1	2	3	4
1	3	2		
2	7	14	1	
3	4	19	42	2
4			1	3

. kap r1 r2 [fw=freq]

Agreement	Expected Agreement	Kappa	Std. Err.	Z	Prob>Z
63.27%	39.65%	0.3913	0.0665	5.88	0.0000

. clear
. *analysis of disagreement (quasi-independence)

. input rr cc y

```
     rr cc  y
  1. 2 1  7
  2. 3 1  4
  3. 4 1  0
  4. 1 2  2
  5. 3 2 19
  6. 4 2  0
  7. 1 3  0
  8. 2 3  1
  9. 4 3  1
 10. 1 4  0
 11. 2 4  0
 12. 3 4  2
 13.end
```

. table rr cc [weight = y]

```
----------------------------------
          |           cc
       rr |    1     2     3     4
----------+-----------------------
        1 |          2
        2 |    7           1
        3 |    4    19           2
        4 |                1
----------------------------------
```

. xi: poisson y i.rr i.cc

```
Poisson regression                                Number of obs   =         12
                                                  LR chi2(6)      =      66.75
                                                  Prob > chi2     =     0.0000
Log likelihood = -15.503339                       Pseudo R2       =     0.6828

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      _Irr_2 |   2.333326   .8817587     2.65   0.008     .6051104    4.061541
      _Irr_3 |   2.367174   .7410715     3.19   0.001     .9147006    3.819648
      _Irr_4 |  -.8850948   1.227905    -0.72   0.471    -3.291744    1.521554
      _Icc_2 |   1.218106   .4555251     2.67   0.007      .325293    2.110918
      _Icc_3 |  -1.103439    .802829    -1.37   0.169    -2.676955    .4700767
      _Icc_4 |  -1.731833   .7696725    -2.25   0.024    -3.240364   -.2233028
       _cons |  -.6651246   .8225334    -0.81   0.419     -2.27726    .9470113
------------------------------------------------------------------------------
```

. predict Y

```
. gen s = sign(y-Y)

. list y Y s

     +--------------------+
     |  y          Y    s |
     |--------------------|
  1. |  7   5.302621    1 |
  2. |  4   5.485178   -1 |
  3. |  0   .2122014   -1 |
  4. |  2   1.738428    1 |
  5. | 19   18.54417    1 |
     |--------------------|
  6. |  0   .7174057   -1 |
  7. |  0   .1705778   -1 |
  8. |  1   1.759029   -1 |
  9. |  1   .0703932    1 |
 10. |  0   .0909944   -1 |
     |--------------------|
 11. |  0   .9383502   -1 |
 12. |  2   .9706554    1 |
     +--------------------+

. egen x2 = sum((y-Y)^2/Y)

. list x2 in 1

     +----------+
     |       x2 |
     |----------|
     | 16.82097 |
     +----------+

. display "pvalue = " 1-chi2(5,x2)
pvalue = .00485214
```

. clear
. *construction of a ROC curve

. infile x class using roc.example.data

. roctab class x, summary table detail

```
           |                           x
     class |     3.294      3.568      3.586      3.725      3.728 |     Total
-----------+-------------------------------------------------------+----------
         0 |         1          1          1          1          1 |        15
         1 |         0          0          0          0          0 |        15
-----------+-------------------------------------------------------+----------
     Total |         1          1          1          1          1 |        30
           |                           x
     class |       3.8      3.808      3.885      3.889      3.916 |     Total
-----------+-------------------------------------------------------+----------
         0 |         1          1          1          1          1 |        15
         1 |         0          0          0          0          0 |        15
-----------+-------------------------------------------------------+----------
     Total |         1          1          1          1          1 |        30
           |                           x
     class |     3.924      3.943      4.003      4.011      4.067 |     Total
-----------+-------------------------------------------------------+----------
         0 |         0          1          0          1          0 |        15
         1 |         1          0          1          0          1 |        15
-----------+-------------------------------------------------------+----------
     Total |         1          1          1          1          1 |        30
           |                           x
     class |     4.071        4.2      4.204      4.212      4.246 |     Total
-----------+-------------------------------------------------------+----------
         0 |         0          1          0          0          0 |        15
         1 |         1          0          1          1          1 |        15
-----------+-------------------------------------------------------+----------
     Total |         1          1          1          1          1 |        30
           |                           x
     class |     4.306      4.349      4.354      4.362      4.375 |     Total
-----------+-------------------------------------------------------+----------
         0 |         1          0          0          1          0 |        15
         1 |         0          1          1          0          1 |        15
-----------+-------------------------------------------------------+----------
     Total |         1          1          1          1          1 |        30
           |                           x
     class |      4.38      4.384      4.411      4.415      4.527 |     Total
-----------+-------------------------------------------------------+----------
         0 |         0          0          0          0          0 |        15
         1 |         1          1          1          1          1 |        15
-----------+-------------------------------------------------------+----------
     Total |         1          1          1          1          1 |        30
```

Detailed report of Sensitivity and Specificity

Cutpoint	Sensitivity	Specificity	Correctly Classified	LR+	LR-
(>= 3.294)	100.00%	0.00%	50.00%	1.0000	
(>= 3.568)	100.00%	6.67%	53.33%	1.0714	0.0000
(>= 3.586)	100.00%	13.33%	56.67%	1.1538	0.0000
(>= 3.725)	100.00%	20.00%	60.00%	1.2500	0.0000
(>= 3.728)	100.00%	26.67%	63.33%	1.3636	0.0000
(>= 3.8)	100.00%	33.33%	66.67%	1.5000	0.0000
(>= 3.808)	100.00%	40.00%	70.00%	1.6667	0.0000
(>= 3.885)	100.00%	46.67%	73.33%	1.8750	0.0000
(>= 3.889)	100.00%	53.33%	76.67%	2.1429	0.0000
(>= 3.916)	100.00%	60.00%	80.00%	2.5000	0.0000
(>= 3.924)	100.00%	66.67%	83.33%	3.0000	0.0000
(>= 3.943)	93.33%	66.67%	80.00%	2.8000	0.1000
(>= 4.003)	93.33%	73.33%	83.33%	3.5000	0.0909
(>= 4.011)	86.67%	73.33%	80.00%	3.2500	0.1818
(>= 4.067)	86.67%	80.00%	83.33%	4.3333	0.1667
(>= 4.071)	80.00%	80.00%	80.00%	4.0000	0.2500
(>= 4.2)	73.33%	80.00%	76.67%	3.6667	0.3333
(>= 4.204)	73.33%	86.67%	80.00%	5.5000	0.3077
(>= 4.212)	66.67%	86.67%	76.67%	5.0000	0.3846
(>= 4.246)	60.00%	86.67%	73.33%	4.5000	0.4615
(>= 4.306)	53.33%	86.67%	70.00%	4.0000	0.5385
(>= 4.349)	53.33%	93.33%	73.33%	8.0000	0.5000
(>= 4.354)	46.67%	93.33%	70.00%	7.0000	0.5714
(>= 4.362)	40.00%	93.33%	66.67%	6.0000	0.6429
(>= 4.375)	40.00%	100.00%	70.00%		0.6000
(>= 4.38)	33.33%	100.00%	66.67%		0.6667
(>= 4.384)	26.67%	100.00%	63.33%		0.7333
(>= 4.411)	20.00%	100.00%	60.00%		0.8000
(>= 4.415)	13.33%	100.00%	56.67%		0.8667
(>= 4.527)	6.67%	100.00%	53.33%		0.9333
(> 4.527)	0.00%	100.00%	50.00%		1.0000

Obs	ROC Area	Std. Err.	-Asymptotic Normal-- [95% Conf. Interval]	
30	0.8978	0.0570	0.78608	1.00000

```
. clear
.* confidence interval for auc

. input  x1  x2
       x1     x2
  1.  3.924 3.294
  2.  4.003 3.568
  3.  4.067 3.586
  4.  4.071 3.725
  5.  4.204 3.728
  6.  4.212 3.800
  7.  4.246 3.808
  8.  4.349 3.885
  9.  4.354 3.889
 10.  4.375 3.916
 11.  4.380 3.943
 12.  4.384 4.011
 13.  4.411 4.200
 14.  4.415 4.306
 15.  4.527 4.362
 16. end

. correlate x1 x2, cov

 (obs=15)

             |       x1       x2
-------------+------------------
          x1 |  .031099
          x2 |  .046358  .080178


. egen xbar1 = mean(x1)
. egen xbar2 = mean(x2)
. gen var1 = _result(3)
. gen var2 = _result(5)
. gen V = (var1+var2)/2
. gen R = (xbar1-xbar2)/sqrt(2*V)
. gen v = (1+R^2/4)/15

. display "lower =", normal(R-1.960*sqrt(v))
  lower = .72301495


. display "upper =", normal(R+1.960*sqrt(v))
  upper = .96137072
```

. clear
. *small example (n = 7)

. input x class

```
         x  class
  1.  24 0
  2.  47 0
  3.  62 0
  4.  12 1
  5.  34 1
  6.  53 1
  7.  68 1
  8. end
```

. roctab class x, summary table detail

```
           |                            x
     class |        12         24         34         47         53 |     Total
-----------+-------------------------------------------------------+----------
         0 |         0          1          0          1          0 |         3
         1 |         1          0          1          0          1 |         4
-----------+-------------------------------------------------------+----------
     Total |         1          1          1          1          1 |         7

           |          x
     class |        62         68 |     Total
-----------+----------------------+----------
         0 |         1          0 |         3
         1 |         0          1 |         4
-----------+----------------------+----------
     Total |         1          1 |         7

Detailed report of Sensitivity and Specificity
                                          Correctly
Cutpoint      Sensitivity   Specificity   Classified          LR+          LR-
------------------------------------------------------------------------------
( >= 12 )         100.00%         0.00%       57.14%       1.0000
( >= 24 )          75.00%         0.00%       42.86%       0.7500
( >= 34 )          75.00%        33.33%       57.14%       1.1250       0.7500
( >= 47 )          50.00%        33.33%       42.86%       0.7500       1.5000
( >= 53 )          50.00%        66.67%       57.14%       1.5000       0.7500
( >= 62 )          25.00%        66.67%       42.86%       0.7500       1.1250
( >= 68 )          25.00%       100.00%       57.14%                    0.7500
( >  68 )           0.00%       100.00%       42.86%                    1.0000
------------------------------------------------------------------------------
                     ROC                    -Asymptotic Normal--
           Obs       Area     Std. Err.      [95% Conf. Interval]
         ------------------------------------------------------------
             7     0.5417       0.2619        0.02841     1.00000
```

. clear
. *Hemoglobin a1c data (long term versus short term)

. input x1 x2

```
        x1   x2
  1.   5.8  5.8
  2.   5.7  4.9
  3.   6.5  6.0
  4.   6.0  5.4
  5.   7.6  7.5
  6.   5.9  5.7
  7.   5.8  5.1
  8.   9.7  9.6
  9.   5.6  5.4
 10.   5.6  4.8
 11.   6.6  6.6
 12.   6.5  5.7
 13.   6.4  5.4
 14.   6.7  6.5
 15.   6.0  6.0
 16.  12.9 12.3
 17.  11.2 10.9
 18.   6.6  6.8
 19.   6.8  5.8
 20.   6.5  5.7
 21.   6.1  5.9
 22.   6.0  5.7
 23.   5.6  5.4
 24.   5.7  5.6
 25.   5.6  5.5
 26.   6.2  6.2
 27.   7.2  7.1
 28.   7.1  7.0
 29.   7.2  7.1
 30.   5.9  5.9
 31.  end
```

. * slope estimated by perpendicular least squares -- ordinary least squares estimation

. regress x1 x2

```
      Source |       SS       df       MS              Number of obs =      30
-------------+------------------------------           F(  1,    28) =  665.55
       Model |  77.8130485     1  77.8130485           Prob > F      =  0.0000
    Residual |  3.27361115    28  .116914684           R-squared     =  0.9596
-------------+------------------------------           Adj R-squared =  0.9582
       Total |  81.0866596    29  2.79609171           Root MSE      =  .34193

------------------------------------------------------------------------------
          x1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x2 |   .9653826   .0374203    25.80   0.000     .8887305    1.042035
       _cons |    .546385   .2490623     2.19   0.037      .036204    1.056566
------------------------------------------------------------------------------
```

```
. correlate x1 x2, cov
             |       x1       x2
-------------+------------------
          x1 |  2.79609
          x2 |  2.77943  2.87909

. gen varx = _result(4)
. gen var1 = _result(3)
. gen var2 = _result(5)
. gen B = ((var1-var2)+sqrt((var1-var2)^2+4*varx^2))/(2*varx)

. list B var1 var2 varx in 1

     +--------------------------------------------+
     |        B       var1       var2        varx |
     |--------------------------------------------|
     | .9851803   2.796092   2.879092   2.779425 |
     +--------------------------------------------+


. *alternative estimate of the least square slope
. gen b = varx/(varx+(var2-varx))

. list b in 1

     +----------+
     |        b |
     |----------|
     | .9653825 |
     +----------+

. *bootstrap estimate of the variance of the slope estimate (perpendicular)

. * slope estimated by perpendicular least squares plus the between and with error variances

. program define slope, rclass
 1.   correlate x1 x2, cov
 2.   gen vx = _result(4)
 3.   gen v1 = _result(3)
 4.   gen v2 = _result(5)
 5.   gen B = ((v1-v2)+sqrt((v1-v2)^2+4*vx^2))/(2*vx)
 6.   summarize B, meanonly
 7.   return scalar B = (mean)
 8.   drop vx v1 v2 B
 9. end
. bootstrap B=r(B), reps(1000): slope

Bootstrap results                               Number of obs      =        30
                                                Replications       =      1000
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           B |   .9851803   .0537582    18.33   0.000     .8798161    1.090544
------------------------------------------------------------------------------
```

. clear
. *Bland/Altman analysis

. input x1 x2

```
          x1    x2
  1.     5.8   5.8
  2.     5.7   4.9
  3.     6.5   6.0
  4.     6.0   5.4
  5.     7.6   7.5
  6.     5.9   5.7
  7.     5.8   5.1
  8.     9.7   9.6
  9.     5.6   5.4
 10.     5.6   4.8
 11.     6.6   6.6
 12.     6.5   5.7
 13.     6.4   5.4
 14.     6.7   6.5
 15.     6.0   6.0
 16.    12.9  12.3
 17.    11.2  10.9
 18.     6.6   6.8
 19.     6.8   5.8
 20.     6.5   5.7
 21.     6.1   5.9
 22.     6.0   5.7
 23.     5.6   5.4
 24.     5.7   5.6
 25.     5.6   5.5
 26.     6.2   6.2
 27.     7.2   7.1
 28.     7.1   7.0
 29.     7.2   7.1
 30.     5.9   5.9
 31. end
```

. gen X = (x1+x2)/2
. gen Y = x1-x2
. regress Y X

```
      Source |       SS       df       MS              Number of obs =      30
-------------+------------------------------           F(  1,    28) =    0.15
       Model |  .017783653     1  .017783653           Prob > F      =  0.7030
    Residual |  3.35588359    28  .119852985           R-squared     =  0.0053
-------------+------------------------------           Adj R-squared = -0.0303
       Total |  3.37366724    29  .116333353           Root MSE      =   .3462

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           X |  -.0147766   .0383608    -0.39   0.703    -.0933551    .0638019
       _cons |   .4209325   .2611379     1.61   0.118    -.1139842    .9558492
------------------------------------------------------------------------------
```

```
. correlate x1 x2,cov
(obs=30)
             |       x1       x2
-------------+------------------
          x1 |  2.79609
          x2 |  2.77943  2.87909

. gen vx = _result(4)
. gen v1 = _result(3)
. gen v2 = _result(5)
. gen V1 = v1-vx
. gen V2 = v2-vx
. egen ybar = mean(Y)
. gen s = sqrt(V1+V2)

. list ybar V1 V2 s in 1

     +-------------------------------------------+
     |     ybar         V1         V2          s |
     |-------------------------------------------|
     | .3233333   .0166667   .0996668   .3410769 |
     +-------------------------------------------+

. gen lower = ybar-1.960*s
. gen upper = ybar+1.960*s

. list lower ybar upper in 1

     +---------------------------------+
     |     lower       ybar      upper |
     |---------------------------------|
     | -.3451775   .3233333   .9918441 |
     +---------------------------------+
```

. clear
. *twins data -- monzygotic
. infile bwt1 bwt2 using twin.mz.data

. correlate bwt1 bwt2, cov

```
(obs=64)
             |     bwt1     bwt2
-------------+------------------
        bwt1 |  .405704
        bwt2 |  .324421  .483741
```

. gen c12 = _result(4)
. gen v1 = _result(3)
. gen v2 = _result(5)
. gen B = ((v2-v1)+sqrt((v1-v2)ˆ2+4*c12ˆ2))/(2*c12)
. egen m1 = mean(bwt1)
. egen m2 = mean(bwt2)
. gen A = m2-B*m1

. list A B v1 v2 c1 in 1

```
     +---------------------------------------------------------+
     |         A          B         v1          v2         c12 |
     |---------------------------------------------------------|
     | -.2133395   1.127478   .4057042   .4837411   .3244213 |
     +---------------------------------------------------------+
```

. reg bwt2 bwt1

```
      Source |       SS       df       MS              Number of obs =      64
-------------+------------------------------           F(  1,    62) =   71.70
       Model |  16.3436732     1  16.3436732           Prob > F      =  0.0000
    Residual |  14.1320186    62  .227935784           R-squared     =  0.5363
-------------+------------------------------           Adj R-squared =  0.5288
       Total |  30.4756918    63  .483741139           Root MSE      =  .47743

------------------------------------------------------------------------------
        bwt2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bwt1 |   .7996498   .0944346     8.47   0.000     .6108777    .9884219
       _cons |   .3008994    .159702     1.88   0.064    -.0183404    .6201392
------------------------------------------------------------------------------
```

. clear
. *dizygotic
. infile bwt1 bwt2 using twin.dz.data

. correlate bwt1 bwt2, cov

```
(obs=67)
             |     bwt1     bwt2
-------------+------------------
        bwt1 |  .631162
        bwt2 |  .516534   .61949
```

. gen c12 = _result(4)
. gen v1 = _result(3)
. gen v2 = _result(5)
. gen B = ((v2-v1)+sqrt((v1-v2)ˆ2+4*c12ˆ2))/(2*c12)
. egen m1 = mean(bwt1)
. egen m2 = mean(bwt2)
. gen A = m2-B*m1

. list A B v1 v2 c1 in 1

```
     +-------------------------------------------------------+
     |        A          B          v1          v2        c12 |
     |-------------------------------------------------------|
     | .0664352   .9887654   .6311617   .6194897   .5165336 |
     +-------------------------------------------------------+
```

. reg bwt2 bwt1

```
      Source |       SS       df       MS              Number of obs =      67
-------------+------------------------------           F(  1,    65) =  139.64
       Model |  27.8997556     1  27.8997556           Prob > F      =  0.0000
    Residual |   12.986563    65  .199793277           R-squared     =  0.6824
-------------+------------------------------           Adj R-squared =  0.6775
       Total |  40.8863186    66  .619489676           Root MSE      =  .44698

------------------------------------------------------------------------------
        bwt2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        bwt1 |   .8183855   .0692545    11.82   0.000     .6800746    .9566963
       _cons |   .3786119   .1381422     2.74   0.008     .1027229     .654501
------------------------------------------------------------------------------
```