Error Matrix for Map Comparison or Accuracy Assessment

These instructions enable you to quantitatively compare two sources of spatial information using a common, non-site specific (i.e. only the total amount of each class/category is measured irrespective of its location) accuracy assessment method. The ***error matrix*** (similar names include confusion matrix, correlation matrix, or covariance matrix) summarizes the relationship between two datasets, often a classification map or model AND reference test information or alternative model. Using ESRI's ArcGIS, Spatial Analyst, and Microsoft's Access and Excel software, perform the following steps to create an error matrix:

- Find all combinations of classes
- Import the data into MS Access
- Create a cross tabulation query to organize the data into a matrix
- Calculate omission error, commission error, overall accuracy, and K_{hat}

ORIGINAL DATA

Map1	raster grid of *n* categories (in this example 5 are used) from a classification model
Map2	raster grid of *n* classes/categories (in this example 5 are used) from an alternative model or comparison reference layer

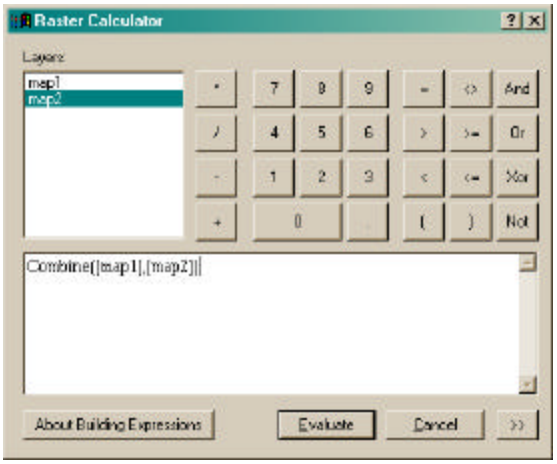
CREATED DATA

Combine	raster grid resulting from the Combine() function
Combine.dbf	dBase table exported from the Combine grid
Error_matrix.mdb	MS Access database to contain the imported combine.dbf table and the created matrix
Combine_Crosstab.xls	MS Excel spreadsheet for calculations

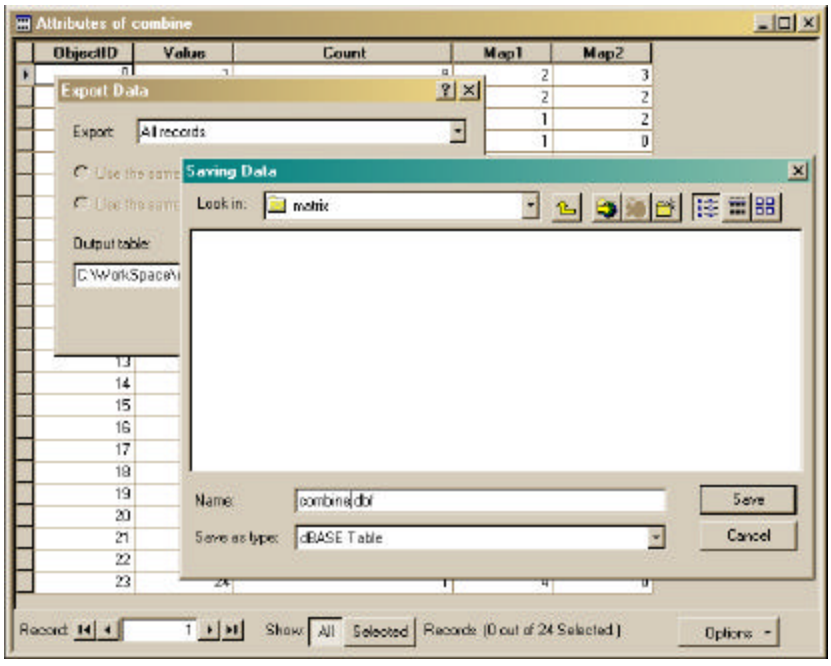
Steps in ArcGIS 8.x:

The following assumes that you have two sources of information that you want to compare ALREADY as raster grids. For help on converting (e.g. tiff, shapefile, etc.) to grid format, please see ArcGIS Desktop Help. Seek help because conversion involves resampling, cell size, and other related concerns that will haunt you!

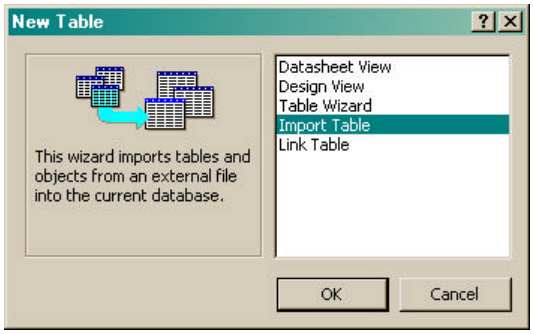
1. Start ArcMap with a new empty map document
2. Choose TOOLS >>> EXTENSIONS
3. Make sure there is a check beside Spatial Analyst – *so it is enabled*
4. Choose VIEW >>> TOOLBARS
5. Make sure there is a check beside Spatial Analyst – *so you can view the toolbar*
6. Click the ADD DATA button and navigate to the directory containing your two layers for comparison
7. Select the appropriate data layers (e.g. **Map1** and **Map2**) and click ADD



8. Choose SPATIAL ANALYST >>> RASTER CALCULATOR
9. Enter the following expression:
   `Combine([Map1], [Map2])`
10. Click EVALUATE
11. Right click on **Calculation** and then click MAKE PERMANENT
12. Save using an appropriate name (e.g. **Combine**)
13. Change the name in the table of contents to **Combine**
14. Right click on **Combine** and OPEN ATTRIBUTE TABLE
15. Click on the OPTIONS button
16. Choose EXPORT
17. Save the file as dBase type using **combine.dbf** as the output name
18. Click SAVE



Interpret the Combine grid table: Map1 field contains the input class value for that layer (similar for Map2). Value assigns a new number to each unique combination of input values, and Count is how many cells share that combination.

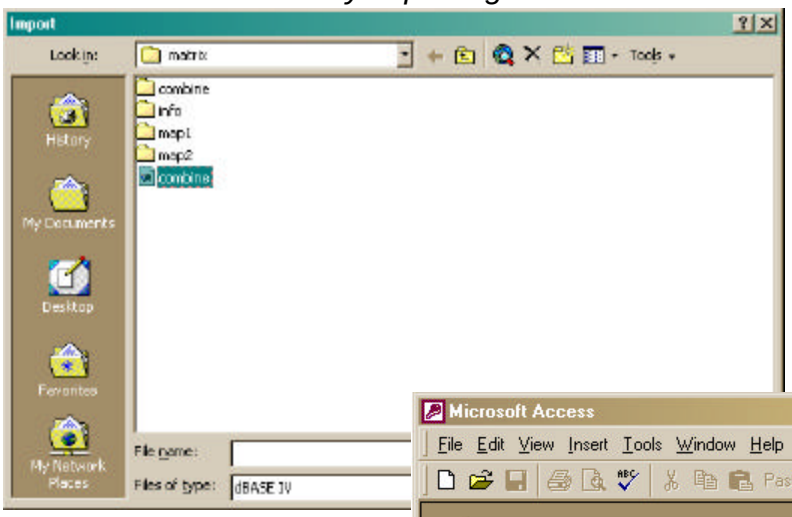


Steps in MS Access:

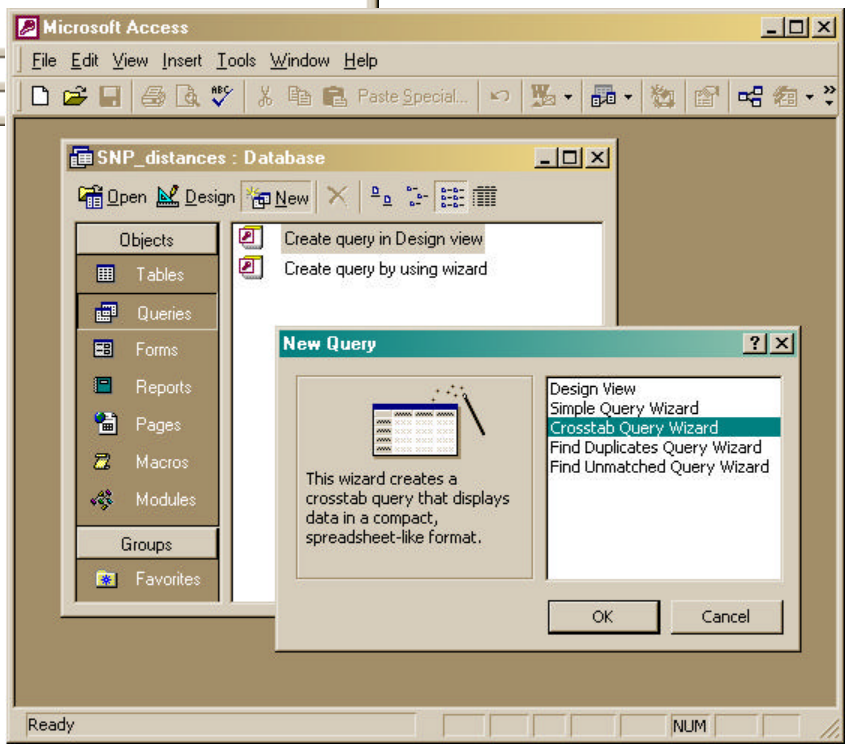
The combination layer that results from the Combine() function simply lists the count of cells for each unique combination of class values between Map1 and Map2. The following assumes prior familiarity with starting an Access database and instructs on how to get the data into a matrix form using this common database program:

1. Start MS Access by creating a new Blank Access database; e.g. **`c:\workspace\matrix\error_matrix.mdb`**

Create a new table by importing the Combine.dbf file.



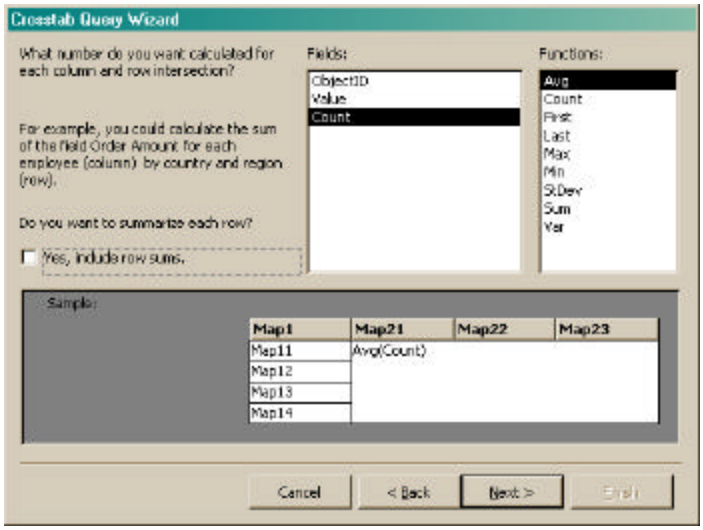
2. Choose INSERT >>> TABLE
3. Select IMPORT TABLE and click OK
4. Navigate to your working directory
5. Select "Files of type:" TEXT FILES
6. Select **Combine.dbf**
7. Click IMPORT



8. Open the new table to view it

It is virtually identical to the attribute table from the Combine raster grid file. Apply a query to create a cross tabulated matrix. A ***crosstab query*** *displays values (and optionally sums, counts, and averages) from one field in a table and groups them by one set of record values listed down the left side of the datasheet and another set of record values listed across the top of the datasheet.*

9. In the Database window, click QUERIES under Objects.
10. Click NEW on the Database window toolbar
11. In the New Query dialog box, click CROSSTAB QUERY WIZARD
12. Click OK
13. Follow the directions in the wizard dialog boxes and click NEXT after each selection:
    - Table = **Combine**
    - Row headings = **Map1**
    - Column headings = **Map2**
    - Specify the field for each column-row intersection = **Count**
    - Specify the field and function = **Avg(Count)**



The default Avg function is okay since the input table contains the single total count for each combination.

14. Click FINISH
15. View the query

This display of the data in matrix format indicates each Map class in the column (Map1) and row (Map2) headings, with the shared number of grid cells between them in the corresponding table cell. If you know your way around MS Access, you may ignore the next section and calculate omission error, commission error, overall accuracy, and K_{hat} statistics (see Jensen, 1996).

combine : Table

ObjectID	Value	Count	Map1	Map2
0	1	9	2	3
1	2	7	2	2
2	3	3	1	2
3	4	6	1	0
4	5	12	0	3
5	6	2	0	4
6	7	11	4	2
7	8	5	0	0
8	9	4	1	4
9	10	8	0	1
10	11	10	3	0
20	21	5	3	3
21	22	5	3	1
22	23	7	3	4
23	24	1	4	0

Record: 1 of 24

combine_Crosstab : Crosstab Query

Map1	0	1	2	3	4
0	5	8	4	12	2
1	6	8	3	7	4
2	5	7	7	9	11
3	10	5	3	5	7
4	1	3	11	7	

Record: 1 of 5

16. Choose FILE >>> EXPORT
17. Save as Microsoft Excel 97-2000

Steps in MS Excel:

There are several references available for quantitatively assessing the correspondence between two raster maps. Remote sensing literature is a good place to look, especially since this is a common method of accuracy assessment on remote-sensing-derived classifications with unbiased ground reference information. Jensen (1996) is the reference applied to this example for testing the relationship between two models.

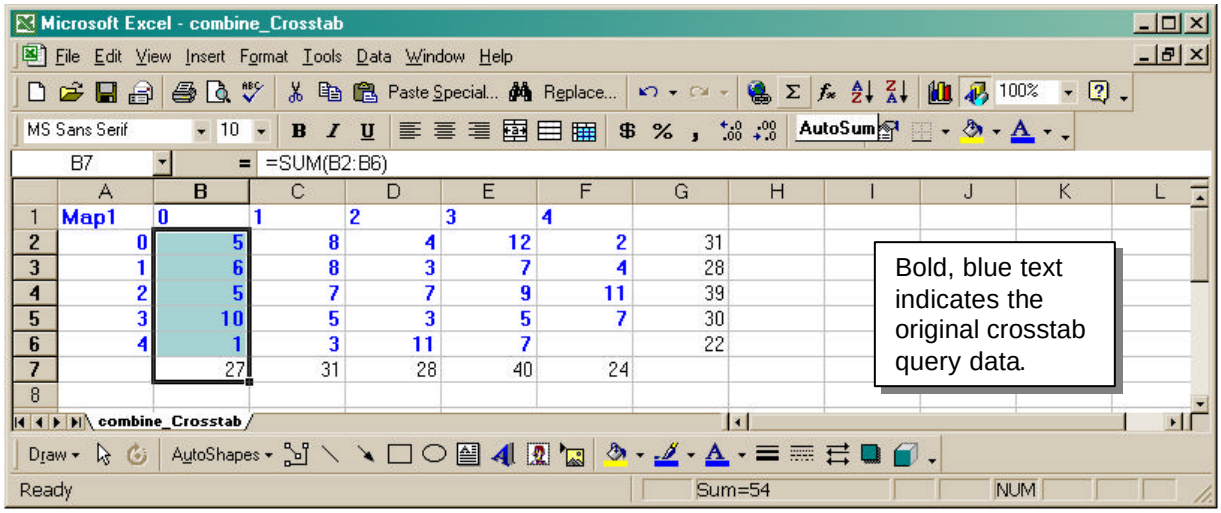
Jensen, John R. 1996. Introductory Digital Image Processing: A remote sensing perspective, 2nd Edition. Prentice Hall: Upper Saddle River, New Jersey. pp. 247-251.

The following assumes prior familiarity with opening an Excel spreadsheet and instructs on how to sum the data and calculate the comparison statistics:

1. Start MS Excel and open the spreadsheet file that you exported from MS Access (i.e. **combine_Crosstab.xls**)

Calculate column totals:

2. Point the cursor in the cell below the first ***column*** of count data (e.g. cell **B7**)
3. Click and drag up to highlight all *values* in the first column
4. Click on the AUTOSUM button



Alternatively, you may enter the formula **=SUM(B2:B6)**.

5. Copy and paste this formula into the second, third, etc. columns

Calculate row totals:

6. Point the cursor in the cell at the end of the first ***row*** or count data
7. Click and drag back to highlight all *values* in the first row
8. Click on the AUTOSUM button
9. Copy and paste this formula into the second, third, etc. rows

Calculate comparison statistics:

10. Read the table below for an overview of the simple descriptive and discrete multivariate statistics involved:

See Jensen, 1996, for a thorough explanation!

Omission Error (a.k.a. producer's accuracy)

Takes into account the accuracy of individual classes; indicates the probability of the cell value in Map2 being the same as in Map1

= x_{ii} / x_{+i} X 100%

x_{ii} = total number correct cells in a class

x_{+i} = sum of cell values in the column

Commission Error (a.k.a. user's accuracy)

Takes into account the accuracy of individual classes; indicates the probability of the cell value in Map1 being the same as in Map2

= x_{ii} / x_{i+} X 100%

x_{ii} = total number correct cells in a class

x_{i+} = sum of cell values in the row

Overall Accuracy

Summarizes the total agreement/disagreement between the maps; only incorporates the major diagonal and excludes the omission and commission errors

= D / N X 100%

D = total number correct cells as summed along the major diagonal

N = total number of cells in the error matrix

K_{hat}

Measure of agreement or accuracy based on KAPPA analysis; useful for comparing maps of similar categories to determine if they are significantly different

$$= N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} X x_{+i}) \Big/ N^2 - \sum_{i=1}^{r} (x_{i+} X x_{+i})$$

r = number of rows in the matrix

x_{ii} = total number correct cells in a class (i.e. value in row *i* and column *i*)

x_{i+} = total for row *i*

x_{+i} = total for column *i*

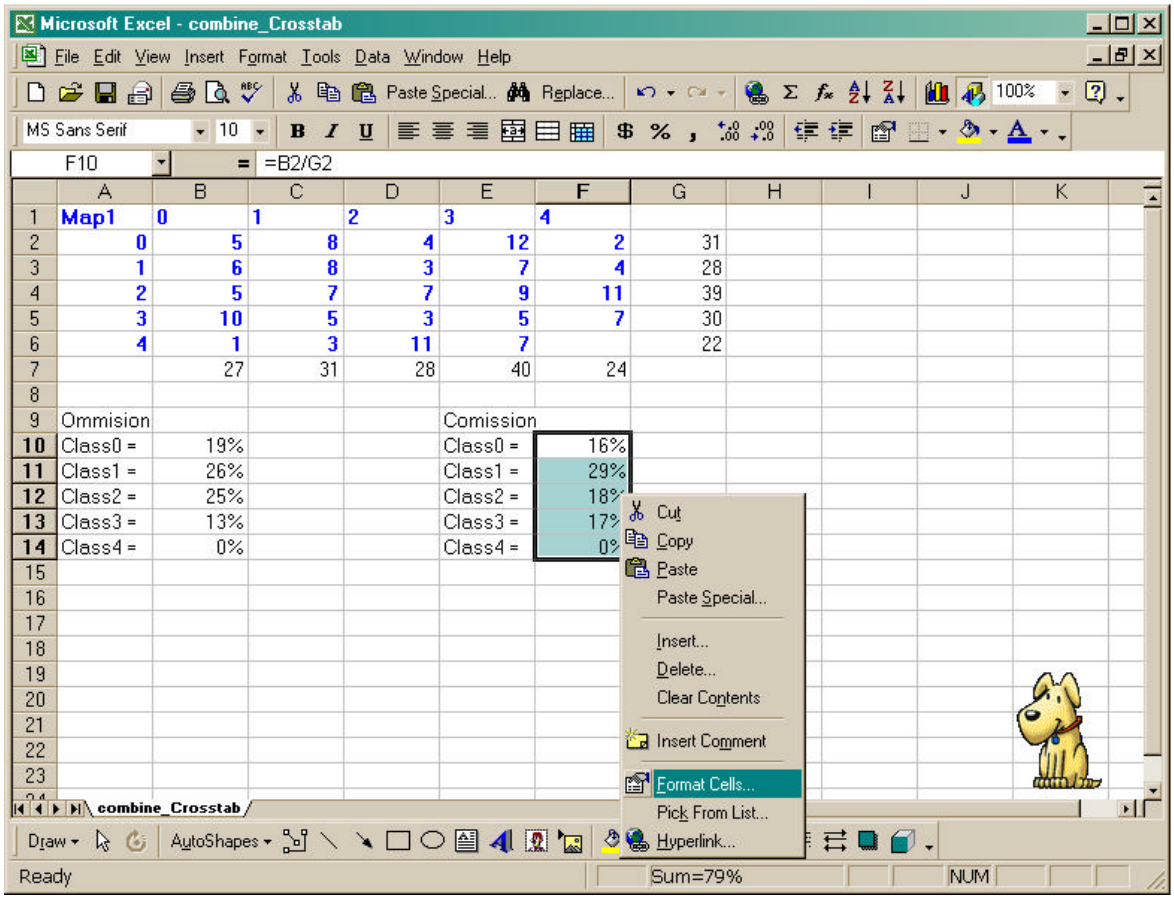
N = total number of cells in the error matrix

For simplicity in calculations (and labeling in MS Excel):

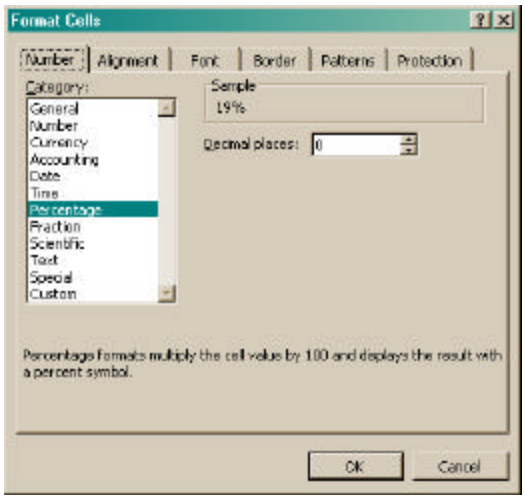
$$\textbf{PART A} = \sum_{i=1}^{r} x_{ii} \qquad \textbf{PART B} = \sum_{i=1}^{r} (x_{i+} X x_{+i})$$

11. Type in the cells as in the figure below to set up the MS Excel Worksheet:
12. Set up labels for each class value under a heading for **Omission** errors

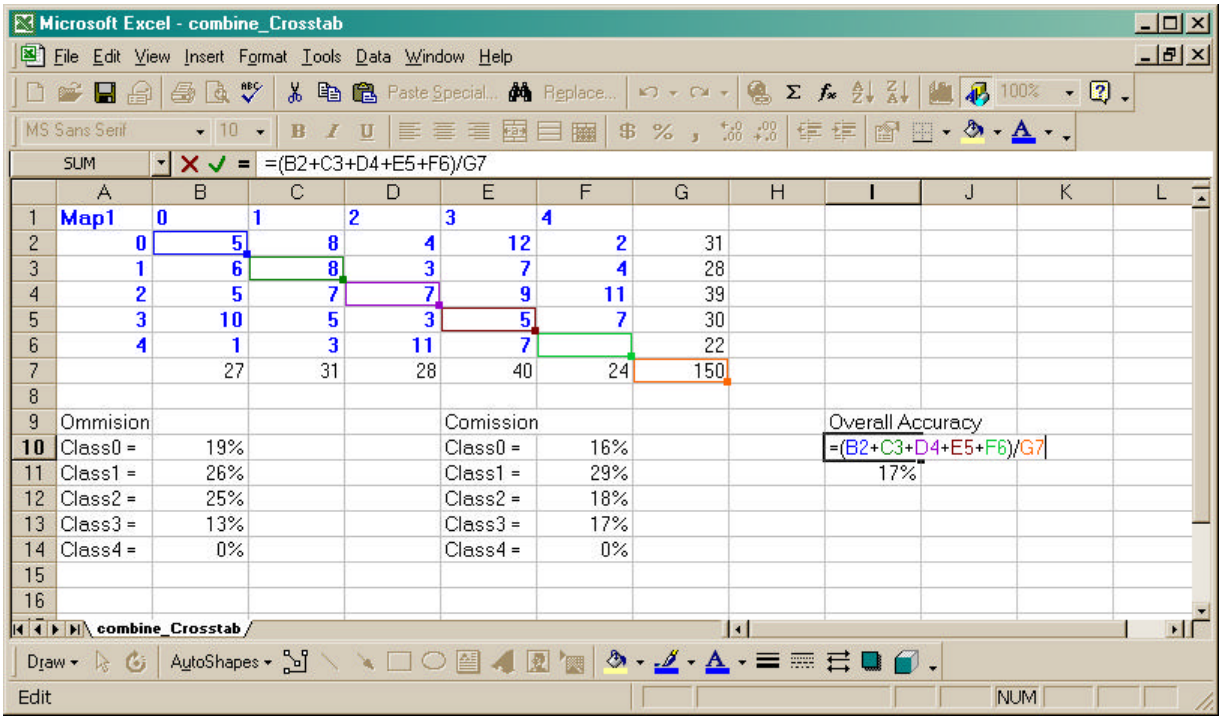
13. Enter the formula **=B2/B7** to calculate the first class' error (i.e. **5/27**)
14. Enter the formula **=C3/C7** to calculate the second class' error (i.e. **8/31**)



15. Continue entering formulae to divide the major diagonal value for a class by the column total
16. Highlight all cells just calculated and RIGHT CLICK
17. Choose FORMAT CELLS
18. Set the number formatting to PERCENTAGE with zero decimal places
19. Click OK
20. Set up labels for each class value under a heading for **Commission** errors
21. Enter the formula **=B2/G2** to calculate the first class' error (i.e. **5/31**)
22. Enter the formula **=C3/G3** to calculate the second class' error (i.e. **8/28**)
23. Continue entering formulae to divide the major diagonal value for a class by the class column total
24. Highlight all cells just calculated and RIGHT CLICK
25. Choose FORMAT CELLS



26. Set the number formatting to PERCENTAGE with zero decimal places
27. Click OK



28. Set up a label for **Overall Accuracy**
29. Enter the formula **=(B2+C3+D4+E5+F6)/G7**
    i.e. ***(5+8+7+5+0)/150***
30. Highlight all the cell just calculated and RIGHT CLICK
31. Choose FORMAT CELLS
32. Set the number formatting to PERCENTAGE with zero decimal places
33. Click OK
34. Set up labels for the $\mathbf{K_{hat}}$ statistic
35. Calculate **N** by summing the row or column totals – *both should be the same!*
36. Enter the formula for **PART A =(B2+C3+D4+E5+F6)**
    i.e. ***5+8+7+5+0***
37. Enter the formula for **PART B**
    =(G2*B7)+(G3*C7)+(G4*D7)+(G5*E7)+(G6*F7)
    i.e. ***(31x27)+(28x31)+(39x28)+(30x40)+(22x24)***
38. Enter the formula for K_{hat} **=(B17*B18)/(B17^2-B19)**
    i.e. ***((N x PART A)-PART B))/(N^2 – PART B)***
39. Set the number formatting for the K_{hat} cell to PERCENTAGE >>> zero decimals

Statistic	Value	MS Excel Formulae
N =	150	**=G7**
PART A =	25	**=(B2+C3+D4+E5+F6)**
PART B =	4525	**=(G2*B7)+(G3*C7)+(G4*D7)+(G5*E7)+(G6*F7)**
Khat =	21%	**=(B17*B18)/(B17^2-B19)**

40. Time to interpret the results! Consult Jensen (1996) or other academic literature