Export Pivot Table to R Using RExcel

Keith Halbert\textsuperscript{1} Richard Heiberger\textsuperscript{1} Erich Neuwirth\textsuperscript{2}

\textsuperscript{1}Department of Statistics, Fox School of Business
Temple University

\textsuperscript{2}Didactic Center for Computer Science and Institute for Scientific Computing
University of Vienna

useR! 2010
Gaithersburg, MD
1. About RExcel
2. About Pivot Tables
3. Why Transfer to R?
4. How it Works
5. Example
6. Conclusion
7. Appendix
Microsoft Excel, available on almost everyone’s machine, provides a familiar interface with which users are comfortable. RExcel is a free add-in to Excel that places the full power of R at the Excel user’s fingertips.

- provides menu access to many R functions directly from Excel by placing the Rcmdr menu on the Excel menu bar
- capable of data transfer to and from R
- allows the placement of R graphs into the Excel automatic recalculation model
- allows any R function to be used in Excel formulas
Once installed, the user connects Excel to R through the Add-ins menu. The help documentation can be accessed from here as well.
Pivot tables are data summarization tools common to spreadsheet software (such as Microsoft Excel). They provide a means for quick and intuitive tabulation of data.

The simple data set and pivot table on the right demonstrates this capability. Here, the average of Height is shown with Sex as a row variable.
The variable selection is made through use of the Field List shown on the right.

Here, Sex has been placed on the rows and we have selected the average of Height to be our summary function.

Excel allows for fields to be dragged, dropped, and filtered in this list.
When Pivot Tables Are Used

- to summarize large transactional data sets
- to find relationships and groupings within data
- to quickly filter data sets
- to organize data in a format that is easy to chart

*Pivot tables are the single most powerful feature in all of Excel.*

— Bill Jelen and Michael Alexander, *Pivot Table Data Crunching*
Why Transfer a Pivot Table to R?

- R graphics are fantastic
- R has many powerful tools for manipulating and analyzing multidimensional tables
- re-creation in R of an existing Excel pivot table could introduce error; automatic transfer is safer
PutPivottable has been included in RExcel since February 2010.

The table created in R is of class structable, a flat contingency table provided in the excellent vcd package (Visualizing Categorical Data).

Options for transfer:
- right-click on a pivot table
- use VBA macro RInterface.PutPivottable (not discussed in these slides)
The Right-Clicking Option

With RExcel activated, the user right-clicks anywhere in a pivot table and selects Put Pivotable.
The previous action brings up the Put table in R dialog box, where the name is assigned. After execution the table may be manipulated as desired using RExcel or the R console.
A Classic Example

The `titanic` data consist of the following variables regarding the passengers of the ill-fated maiden voyage of the Titanic:

- **Class**: levels first, second, third and crew
- **Age**: levels child and adult
- **Sex**: levels M and F
- **Survive**: levels die and live

These are recoded from the original. See the Reference section for more information. Two of the 2201 records follow.

<table>
<thead>
<tr>
<th>Class</th>
<th>Age</th>
<th>Sex</th>
<th>Survive</th>
</tr>
</thead>
<tbody>
<tr>
<td>third</td>
<td>child</td>
<td>F</td>
<td>die</td>
</tr>
<tr>
<td>crew</td>
<td>adult</td>
<td>M</td>
<td>live</td>
</tr>
</tbody>
</table>
Titanic Example

We begin with a simple statistical question: Does the class of the passenger help to explain survival?

We place Class on the rows and Survive on the Columns.

<table>
<thead>
<tr>
<th>Class</th>
<th>Survive</th>
<th>die</th>
<th>live</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>crew</td>
<td></td>
<td>673</td>
<td>212</td>
<td>885</td>
</tr>
<tr>
<td>first</td>
<td></td>
<td>122</td>
<td>203</td>
<td>325</td>
</tr>
<tr>
<td>second</td>
<td></td>
<td>167</td>
<td>118</td>
<td>285</td>
</tr>
<tr>
<td>third</td>
<td></td>
<td>528</td>
<td>178</td>
<td>706</td>
</tr>
<tr>
<td>Grand Total</td>
<td></td>
<td>1490</td>
<td>711</td>
<td>2201</td>
</tr>
</tbody>
</table>
To perform a chi-square test of independence, Excel requires us to manually create a table of expected cell counts. These are the numbers we would expect if Class and Survive were independent.

\[
\text{CHITEST()} \text{ calculates a p-value (which indicates dependence).}
\]
Titanic Example

The chi-square test is an approximation. For some count data we may require Fisher’s exact test, which Excel does not provide. To perform either of these tests in \( R \), first we transfer the table and give it a sensible name.
Titanic Example

The RExcel function `REval()` evaluates R code and brings the result to Excel. The chi-square approximation is fine for these data. The p-values are similarly tiny.

In `chisq.test()` R calculates the table of expected values directly from the pivot table. The manual calculation is not necessary.

<table>
<thead>
<tr>
<th>result</th>
<th>formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>chi-square in R</td>
<td>=REval(&quot;chisq.test(titanic.table)$p.value&quot;)</td>
</tr>
<tr>
<td>Fisher Exact in R</td>
<td>=REval(&quot;fisher.test(titanic.table, workspace=1000000)$p.value&quot;)</td>
</tr>
</tbody>
</table>

Even in this simple example, Excel cannot compete with R when it comes to statistical analysis. We now turn our attention to graphics.
Titanic Example

Below is a standard Excel plot of our pivot table.

![Titanic Example Chart](chart.png)

This is a simple table and Excel’s plot is OK. Next we examine a mosaic plot from package `vcd`. These are great for comparing count data, but are not available in Excel.
Mosaic Plot

- Survive
- Die
- Class
- Crew
- First
- Second
- Third

Crew: 2/3, 1/3
First: 1/3, 1/3
Second: 1/3, 1/3
Third: 1/3, 2/3
Titanic Example

We add two additional fields: a second row variable (Sex) and a second column variable (Age).

<table>
<thead>
<tr>
<th>Class</th>
<th>Sex</th>
<th>Age</th>
<th>Survive</th>
<th>Die</th>
<th>Child</th>
<th>Survive</th>
<th>Die</th>
<th>Child</th>
<th>Survive</th>
<th>Die</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>crew</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td>670</td>
<td></td>
<td></td>
<td>192</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>first</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td>118</td>
<td></td>
<td></td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>second</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td></td>
<td></td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td>154</td>
<td></td>
<td></td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>third</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td>89</td>
<td></td>
<td></td>
<td>76</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td>387</td>
<td></td>
<td></td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Grand Total</strong></td>
<td></td>
<td></td>
<td></td>
<td>1438</td>
<td></td>
<td></td>
<td>654</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


Titanic Example

Below is the best we can do in Excel. The mosaic plot on the next slide is superior. Excel pivot table users should have access to these plots.
The Mosaic Plot is Superior
Conclusion

- RExcel allows for instantaneous transfer of pivot tables from Excel to R.
- Once the table is transferred, the user has the full power of R to do statistical analysis.
- R packages such as vcd provide useful graphics that are not available in Excel.
References and Links

Baier, T. and Neuwirth, E. (2007), Excel :: COM :: R.
  Computational Statistics 22/1 91-108

  Journal of Statistics Education v.3, n.3

Heiberger, R. and Neuwirth, E. (2009), R Through Excel. Springer Verlag

Hornik, K. The Strucplot Framework: Visualizing Multi-way Contingency Tables
  vol. 17(i03).

Jelen B. and Alexander, M. (2005), Pivot Table Data Crunching. Que Publishing

Report on the Loss of the 'Titanic' (S.S.) (1990),
  British Board of Trade Inquiry Report (reprint). Allan Sutton Publishing

RExcelInstaller at CRAN
  http://cran.r-project.org/web/packages/RExcelInstaller/index.html

RExcel Homepage
  http://rcom.univie.ac.at/

Titanic Dataset
  http://www.amstat.org/publications/jse/v3n3/datasets.dawson.html
R Code for Creating Mosaic Plots

mosaic(~ Class + Survive,
       direction = c("v", "h"),
       data = test,
       highlighting = "Survive",
       highlighting_fill = c("salmon", "skyblue")
   )

mosaic(~ Class + Survive + Sex + Age,
       direction = c("v", "h"),
       data = test,
       highlighting = "Survive",
       highlighting_fill = c("salmon", "skyblue")
   )