Some Advice on Tables and Figures

This short guide will hopefully help you avoid common mistakes while preparing tables and figures for your paper. The golden rule is that tables and figures should be self-contained. This means that the reader should be able to pretty much figure out what is in the table or figure without having to look out other tables or reading an extensive description of the table (or figure) in the text. Here’s a particular bad example of a table that fails to do so:

Table 1: Regression results

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED1</td>
<td>-0.1023425***</td>
<td>-0.02222232*</td>
<td>-0.15252525**</td>
</tr>
<tr>
<td></td>
<td>(-4.4372894)</td>
<td>(-1.90435454)</td>
<td>(-2.380543789)</td>
</tr>
<tr>
<td>ED3</td>
<td>0.425543425***</td>
<td>0.384545532***</td>
<td>0.554354525***</td>
</tr>
<tr>
<td></td>
<td>(8.5435894)</td>
<td>(6.54354354)</td>
<td>(4.3425589)</td>
</tr>
</tbody>
</table>

Note: T-statistics in parentheses. "*", "**", "***" indicates significance level of 90, 95, and 99 percent, respectively.

(I have not reported results for the “IND” variables but similar numbers should be there too)
Contrast this with a much more informative version of the same table:

Table 1: Estimates of the Log-wage Equation for Canadian Women: a Comparison of OLS and Instrumental Variables (IV) Results

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school dropouts</td>
<td>-0.102***</td>
<td>-0.022*</td>
<td>-0.153**</td>
</tr>
<tr>
<td>(0.024)</td>
<td>(0.013)</td>
<td>(0.065)</td>
<td></td>
</tr>
<tr>
<td>University graduates</td>
<td>0.426***</td>
<td>0.385***</td>
<td>0.554***</td>
</tr>
<tr>
<td>(0.054)</td>
<td>(0.060)</td>
<td>(0.139)</td>
<td></td>
</tr>
<tr>
<td>Industry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manufacturing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupation dummies?</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Standard errors in parentheses. "*", "**", "***" indicates significance level of 90, 95, and 99 percent, respectively. In column 3, the two education variables (dropout and university graduates) are instrumented using compulsory schooling laws. All three models are estimated using 2,343 observations. All models also include province dummies. See text for more detail on the sample.
This new table is self contained and much more informative than the first table for a number of reasons like

1. We know from the title what model we are estimating for which sample.
2. We know what is the difference between the three models without having to look up the text to figure out what “Model 1”, “Model 2” and “Model 3” are.
3. The acronyms for the variables (ED1, etc.) have been replaced by informative headings such as “high school dropouts”.
4. We gained space to include informative variable headings in part because we have removed unnecessary digits at from the estimated coefficient.
5. We report standard errors instead of t-statistics underneath the estimated coefficients. Since we already have “stars” for significance, t-statistics are fairly useless as they restate the same information (e.g. you get two stars when the t-stat is above 2). The standard errors add more information as they give a much better sense of confidence intervals. This also indicates why adding more digits would be useless. For example, the very first standard error of 0.024 tells you that it is meaningless to add digits beyond the third digit.

Just cutting and pasting from a computer output is not a good way of designing a nice and informative table. You have to decide what is important to report (here we decided that reporting occupation and province dummies did not add much) and make sure that all the information the reader needs is there.

The same basic rules apply to figures. Make sure you have an informative title and don’t hesitate to add detailed notes to explain what is in the figure. Make sure you don’t present too much information in the figure, so that the eye can see what is going on. Smoothing is often useful to help illustrate strong empirical regularities as opposed to statistical noise.