1  Downloading files and accessing SAS.

We will be using the “billion.dat” dataset again today, as well as the OECD dataset on health care expenses. Read the file “OECD.info” to learn about the OECD dataset. Also look at the OECD.data dataset.

Then call up SAS.

2  Sorting, scatterplots, correlation and regression

In the following SAS code, lines that begin with an asterisk are comments and do not need to be typed.

```
************************************
* Setting the number of characters *
* in output lines and pages *
************************************
options linesize = 79 pagesize = 60;

*******************
* Reading the billionaire *
* dataset into SAS *
*******************
* Use this version if you are running SAS on the computer you are on ;
data billion;
infile 'c:\temp\billion.dat';
input with age region $;
run;

* Use this version if you are running SAS on the Virtual Desktop ;
data billion;
infile with age region $;
datalines;
<paste data in here >
run;

********************
* Sorting a dataset *
********************
* Note: If we want to produce separate output for different subsets of *
a dataset, we must first sort the dataset by the variable that *
defines those subsets ;
proc sort data = billion;
by region;
```

```
run;

***********************************
* Producing separate analysis for *
* each region *
***********************************
* Note: In addition to a complete univariate analysis within each *
* region, this procedure produces side-by-side boxplots of wealth *
* by region ;
proc univariate plot data = billion;
var with;
by region;
run;

***************************
* Producing a scatterplot *
***************************
* Note: the following code plots with on the y-axis and age on the x-axis;
proc plot data = billion;
plot with * age;
run;

*************************
* Reading the OECD dataset *
* into SAS *
*************************
* Note: the "13." in the "input" statement tells SAS the number of *
* characters in the longest country name. Without this information, *
* SAS would truncate the country names to 8 letters each ;
data OECD;
infile country $ 13. pcgdp pch beds los docs infmort;
datalines;
<paste data here>;
run;

*****************************
* Better text scatter plots *
*****************************
proc plot data = OECD;
plot pch * pcgdp = '.' / vpos = 20 hpos = 40;
run;

***************
* Correlation *
***************
```

```
proc corr data = billion;
run;
```

***************
* Correlation *
***************
```
**Analyst for regression**

Use the following steps to get into “Analyst” from the menu:

- **Solutions**
  - Analysis
    - Analyst

You must specify which dataset you wish to use. Do so by clicking

---

**Insight for regression**

Insight is another point-and-click facility built into SAS. We will be using its graphical features later on when we study multiple regression. In case you want to try it now, here are some instructions.

From the main pull-down window, select the following sequence of choices:

- **Solutions**
  - Analysis
    - Interactive data analysis

In the window that appears, you must specify which dataset you wish to use. Do so by clicking

- **Library**: Work
  - **Dataset**: OECD
    - Open

To do regression in Insight, choose

- **Analyze**
  - Fit

To identify the response variable, use your mouse to click “PCH” and then “Y.” Similarly, copy “PCGDP” into the “X” column. Click “OK” and lots of regression output and plots will appear.

To get out of Insight and back into command mode, click in the window showing the data in spreadsheet form. Then pull down the “File” menu and choose “End.”

---

**Remember to exit from SAS and log out of your hawkid**
### MODEL 1
**Dependent Variable:** PCH

#### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Squares</th>
<th>Square</th>
<th>F Value</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>12390694.827</td>
<td>12390694.827</td>
<td>87.518</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>27</td>
<td>3822637.863</td>
<td>141579.180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C Total</td>
<td>28</td>
<td>16213332.69</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Root MSE**: 376.27009
- **R-square**: 0.7642
- **Dep Mean**: 1508.89655
- **Adj R-sq**: 0.7555
- **C.V.**: 24.93677

#### Parameter Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>T for H0:</th>
<th>Prob &gt;</th>
<th>T</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEP</td>
<td>1</td>
<td>-465.663682</td>
<td>222.33243595</td>
<td>-2.094</td>
<td>0.0457</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCGDP</td>
<td>1</td>
<td>0.096818</td>
<td>0.01034925</td>
<td>9.355</td>
<td>0.0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### MODEL 2
**Dependent Variable:** PCH

#### Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Squares</th>
<th>Square</th>
<th>F Value</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>12390694.827</td>
<td>12390694.827</td>
<td>87.518</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>27</td>
<td>3822637.863</td>
<td>141579.180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C Total</td>
<td>28</td>
<td>16213332.69</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Root MSE**: 376.27009
- **R-square**: 0.7642
- **Dep Mean**: 1508.89655
- **Adj R-sq**: 0.7555
- **C.V.**: 24.93677

#### Parameter Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>T for H0:</th>
<th>Prob &gt;</th>
<th>T</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEP</td>
<td>1</td>
<td>-465.663682</td>
<td>222.33243595</td>
<td>-2.094</td>
<td>0.0457</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCGDP</td>
<td>1</td>
<td>0.096818</td>
<td>0.01034925</td>
<td>9.355</td>
<td>0.0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Obs COUNTRY PCH Predict

<table>
<thead>
<tr>
<th>Obs</th>
<th>COUNTRY</th>
<th>PCH</th>
<th>Predict</th>
<th>Value</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Australia</td>
<td>1775.0</td>
<td>1731.0</td>
<td>43.9558</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Austria</td>
<td>1748.0</td>
<td>1856.5</td>
<td>-108.5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Belgium</td>
<td>1708.0</td>
<td>1867.4</td>
<td>-169.4</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Canada</td>
<td>2065.0</td>
<td>1903.3</td>
<td>161.7</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>CzechRepub</td>
<td>904.0</td>
<td>806.2</td>
<td>97.7631</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Denmark</td>
<td>1802.0</td>
<td>2078.7</td>
<td>-276.7</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Finland</td>
<td>1380.0</td>
<td>1631.3</td>
<td>-251.3</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>France</td>
<td>2002.0</td>
<td>1673.1</td>
<td>328.9</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Germany</td>
<td>2278.0</td>
<td>1745.2</td>
<td>532.8</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Greece</td>
<td>888.0</td>
<td>934.6</td>
<td>-46.6178</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Hungary</td>
<td>602.0</td>
<td>553.3</td>
<td>48.7491</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Iceland</td>
<td>1893.0</td>
<td>2080.3</td>
<td>-187.3</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Ireland</td>
<td>1276.0</td>
<td>1713.6</td>
<td>-437.6</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Italy</td>
<td>1584.0</td>
<td>1639.1</td>
<td>-55.0669</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Japan</td>
<td>1677.0</td>
<td>1868.5</td>
<td>-191.5</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Korea</td>
<td>537.0</td>
<td>845.3</td>
<td>-308.3</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Luxembourg</td>
<td>2139.0</td>
<td>2878.0</td>
<td>-739.0</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Mexico</td>
<td>368.0</td>
<td>308.7</td>
<td>331.4</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Netherlands</td>
<td>1766.0</td>
<td>1769.1</td>
<td>-3.0938</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>NewZealand</td>
<td>1270.0</td>
<td>1249.2</td>
<td>20.8199</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Norway</td>
<td>1928.0</td>
<td>2196.5</td>
<td>-268.5</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Poland</td>
<td>371.0</td>
<td>307.5</td>
<td>63.4736</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Portugal</td>
<td>1071.0</td>
<td>1012.4</td>
<td>58.6372</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Spain</td>
<td>1115.0</td>
<td>1156.1</td>
<td>-60.0728</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Sweden</td>
<td>1675.0</td>
<td>1888.1</td>
<td>86.8946</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Switzerland</td>
<td>2499.0</td>
<td>2108.3</td>
<td>390.7</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Turkey</td>
<td>232.0</td>
<td>185.0</td>
<td>47.0454</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>UnitedKingdom</td>
<td>1317.0</td>
<td>1584.0</td>
<td>-267.0</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>UnitedStates</td>
<td>3898.0</td>
<td>2488.6</td>
<td>1409.4</td>
<td></td>
</tr>
</tbody>
</table>

9

10