Basic Data Analysis

Stephen Turnbull
Business Administration and Public Policy

Homework 5: Due June 9, 2011

Requirements

Homework assignment about creating a data set.

Homework 5

Due Thursday, June 9, 2011, 11:45 am. Submit by email to `data-hw@turnbull.sk.tsukuba.ac.jp`. Your header should look like this:

```
From:  a-student@sk.tsukuba.ac.jp
To:  data-hw@turnbull.sk.tsukuba.ac.jp
Subject:  Basic Data Analysis HW#5
```

The subject should be all half-width Roman letters (ASCII).

Create a dataset

1. **Find data** for **three** variables which are related to each other. You must provide at least 20 values for each variable. *Briefly* describe your variables and their relationship. *The credit you receive for this problem will be inversely proportional to the number of other students who choose the same example.*

   These variables may be related to any field of interest to you, not necessarily limited to business.

2. **Choose a statistical package**. I will be using **R** (`http://www.r-project.org/`) because it is *open source* (you can get it free from the R Project) and I use it myself. Other packages easily available in Keisei include **S+** (very similar to R with better GUI), **SPSS**, **TSP**, and also I believe **SAS** and **Shazam** are available. There may be others, all are sufficient for our purposes. *Warning*: **Excel** and other spreadsheet programs are *not* acceptable as statistical packages.

   Which package did you choose? Why? (I'm just curious, you will not lose points on "why?" as long as you answer the question.)

   My advice: **R** and **S+** are highly programmable research tools. Since I'm using R, you will not have trouble with them in this class. However, in the future you may find them too complex. **TSP** and **Shazam** are oriented toward time series analysis, and regression analysis in general. Students interested in finance, economics, or demand analysis in marketing may find these most useful in the future. **SPSS** may be the most GUI-oriented, and technically it favors cross-section analysis, and ANOVA. Students interested in product design aspects of marketing and organizational behavior may find SPSS most useful. **SAS** is a very powerful data manager as well as statistical analysis system.

3. **Create a dataset** from the variables using your package. If you use a *program* to do this, please either include it here or attach it as a file. If you used the GUI, explain the procedure.

4. **Attach the data file** used to create your dataset, if you used one.

5. **Print your dataset to a file** using your statistical package. Attach the file.