

# Basic Data Analysis

Stephen Turnbull

Business Administration and Public Policy

Homework 2: Due May 12, 2011

## Requirements

Homework assignment on distributions in Excel.

## Homework 2

**Due 2011-05-12, 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#2
```

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

1. Download `homework-sheet-2.xls` from the home page. You also may want to get `mean-variance.xls` (used in class).
2. Copy the random numbers (values only!) to the *original data set* and construct the *sorted data set* as done in class.
3. Construct the absolute and relative frequency distributions, and the cumulative (relative) frequency distribution, of *letter grades*. (Space in the worksheet is allocated for the *dummy variable* technique shown in class, but if you prefer you can omit the dummy variables and use a count function or distribution function provided by the spreadsheet software.)
4. Translate the letter grades to the usual 4-point scale.

5. Enter the *mode* and the *median* in the appropriate places.
6. Enter formulæ to compute the various components of the *mean*, *variance*, and *standard deviation*. Also compute the variance using the “moment difference” method, and compare to the “central moment” method. They should be exactly the same.
7. In some empty space, draw a histogram of your distribution by using colored cells as shown in class.
8. Answer the questions following the distribution exercise in the spaces provided.

**Important notes:** For your convenience, areas provided for your answers are highlighted in blue.

In answering the questions, *you must explain your answers in some detail*. For numerical problems, you may provide the basic formulæ and detailed algebraic calculations (in general you should not show numerical calculations unless explicitly requested). For questions testing definitions, you should provide the definition and explain how the example corresponds to the conditions of the definition. For questions testing interpretation, you should describe the logic that leads to your conclusion.

There are a lot of questions on *skewness* and *kurtosis* in the problem set. You should not take this as an indication of the importance of these statistics themselves. Instead, the exercises are intended to give you some “feeling” for the relationship between summary statistics and the “shape” of the distribution that they describe.

There are only two pure “interpretation” questions, no. 8 and no. 9. However, these are very important, and similar questions will be featured in both the midterm and the final examinations.