WST211

Practical 1

Date: 8 February 2016
Homework – Hand in at the latest on 15 February 2016 @ 13h30

Contents:
The practical activities are split into two sections:

Section A: This is intended as a revision exercise of the SAS programming activities from the first year level. It is slightly more comprehensive, but requires the same procedures as used at the first year level.

Section B: Simulations in SAS. The content is aimed at enhancing theoretical aspects, usually through simulation studies.

Instructions:

- Hand in a typed document with your answers which includes the following:
  - Table of contents
  - Answers
  - SAS program
  - SAS output
- Attach a copy of the SAS program in an appendix.
- Attach the relevant outputs of the SAS output.
- According to the question, make interpretations about the SAS output.
- Round the answers to 3 decimal places.

Section A
Water quality is determined by measuring particulate levels of several substances. The Utah Division of Water Quality collected water samples from a specific well for the period January 1990 to December 1991 and measured the nitrate, zinc and TDS (total dissolved solids) levels for each sample.
The dataset **well.dat** is given on ClickUP.

Variable	**Columns**
Date (month/day)	1-5
Year	7-8
Nitrate (mg/l)	11-15
Zinc (mg/l)	18-22
TDS (mg/l)	25-27

.

The dataset is space delimited and can be imported into SAS by using the following statements:

```
data well;
infile 'c:\well.dat' dlm=' ';
input v1$ v2 v3 v4 v5;
```

Questions

1. Give the mean and standard deviation of the nitrate, zinc and TDS measurements for all the samples collected.
2. Draw joint boxplots in SAS to compare the zinc of samples collected in 1990 and in 1991. Hand in the relevant SAS output.
3. Give the five-point summary statistics for zinc of samples collected in 1990 and 1991 by completing the following table.

Five-point summary of zinc for 1990 and 1991

Summary statistic	1990	1991
Minimum		
First quartile		
Median		
Third quartile		
Maximum		

Section B

Study the notes "Simulations in SAS" (available on ClickUP) and answer the questions that follow.

Questions

1. Draw 10000 values randomly from a standard normal population and determine the following probabilities empirically (that is through a simulation). Use a seed value of 50.
   a) P (Z > 1.8), Z ~ N(0,1).
   b) P (–1.024 < Z < 1.335), Z ~ N(0,1).

   Note: The rannor(seed) function generates a value randomly from the N(0,1) population. The seed value can be any integer. If 0 is chosen, the values generated will differ for each simulation.

2. A man pays R10 to win a R30 Kewpie doll. His probability of winning on each throw is 0.2. He must win a doll for each of his three children. Let X be the random variable indicating the number of throws required to win three dolls. Determine through simulation empirical values for the following. (Use 1000000 simulations and a seed value of 1.)
   a) The probability that 10 throws are needed to win the four dolls, that is P (X = 10).
   b) The probability that at least ten throws will be required, that is P (X ≥ 10).
   c) The probability that at least seven throws but no more than eighteen throws will be required, that is P (7 ≤ X < 18).
   d) The expected number of throws needed to win four dolls, that is $E(X)$.

   Note: The ranuni(seed) function generates a value randomly from the UNIF(0,1) population. In this problem, a value in the interval [0,0.3] can be considered as winning on a throw.