**STAT 40001/MA 59800 Statistical Computing/ Computational Statistics Fall 2013**

**Test 1- Form II**

**Name:**

**PUID:**

*This exam consists of 5 questions of worth 100 points. Please provide the R codes that have been used to perform the calculations and graphics along with the interpretation of the output. It is important that you distinguish between the R code and the description. You may use different fonts or different colors.*

**Q.N. 1)** **Short answer questions**

a) Create a vector named countby5 that is a sequence of 5 to 100 in steps of 5.

b) Provide R code to create the sequence 5, 10, 10, 15, 15, 15, 20, 20, 20, 20.

c) Generate 100 random numbers from a normal distribution with mean 10 and variance 25. Please print first 5 observations.

d) Generate 50 random numbers form a t- distribution with 18 degrees of freedoms. Please print first 5 observations.

e) The brightness dataset in the UsingR package contains the information about the brightness of stars in a sector of the sky. How many observations are included in the dataset? Please print first 5 observations.

**Q.N. 2)** Seven subjects were assigned to group A and eight subjects were assigned to group B. The two groups using different training methods to improve the subjects’ read speed. The results are as following:

Group A: 500 700 250 404 390 555 589

Group B: 355 388 445 469 560 502 430 480

Read the data above to the R system. Perform t-test and a nonparametric test to compare the two groups A and B. Check if it is appropriate to use t-test. Compare the results of t-test and nonparametric test.

**Q.N. 3)** The chickwts data are collected from an experiment to compare the effectiveness of various feed supplements on the growth rate of chickens and are available in the base package.

a) How many variables are in the database?

b) Display the information by creating side-by-side boxplot by choosing an appropriate variable. Please make sure that you have appropriately labeled the axes and use suitable title of the boxplot.

**Q.N. 4)** The National Restaurant Association is interested in determining if there is a relationship between the types of pizza pie which Americans prefer and the region of the country in which they live . The association randomly selects 285 Americans and records the category of pizza pie which best describes their preference and the region of the country they live. The data are recorded as below

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type of  Pizza Pie  Preferred |  | North | South | East | West |
| Thin Crust | 40 | 30 | 35 | 45 |
| Thick Crust | 17 | 15 | 21 | 22 |
| Pan Pizza | 15 | 15 | 15 | 15 |

Can the association conclude that the types of pizza pie which Americans prefer and the region of the country they live are dependent at 0.01?

**Q.N. 5)** Journal of Statistics Education, Volume 4, Number 2 (July 1996) include an article What's Normal? -- Temperature, Gender, and Heart Rate by A. Shoemaker. The dataset used in the article are provided in <http://www.amstat.org/publications/jse/datasets/normtemp.dat.txt>. The description of the data can be accessed in the link below.

<http://www.amstat.org/publications/jse/datasets/normtemp.txt>

1. How many variables are included in the study?
2. Print first five observations of the data.
3. Is the distribution of body temperatures normal?
4. Is the true population mean really 98.6 degrees F?
5. Is there a significant difference between males and females average temperature?