STAT:5400 Computing in Statistics
Fall 2015
Final Exam

You may submit your answers either as:

- a plain text file:
  Number the questions. Include all R and SAS code and R output, and put your name as the first line of the file.

- or, for 2 percentage points extra credit, as an .Rnw file and PDF file:
  embed R code into the .Rnw file, and just copy and paste SAS code into a `verbatim` environment.

SAS output may either

- be included as part of your text file, or in a verbatim environment of your .Rnw file, or

- be submitted in a separate PDF file

Run R (and possibly LATEX) on the Linux machines. You may not use your own laptop. You may run SAS either under Linux or on the Virtual Desktop.

1   Bootstrap

A dataset called `problem2.dat` is available under "Datasets" on the course web page.

1. Read this dataset into a vector object in R.

2. Perform a nonparametric bootstrap of the interquartile range statistic.

3. Perform a parametric bootstap of the interquartile range statistic under the assumption that the data is a sample from a normal population.

4. Comment on which estimate of bias is likely to be more correct and why.

2   SAS

Two datasets, `patients2015.dat` and `patientVisits2015.dat` are on the course web page under Datasets. Documentation is provided in corresponding files with the extension `.info`.

1. Read both datasets into SAS, and merge them in such a way that only patients who have visits are included in the combined dataset. Display the combined dataset. Submit all your SAS code and the output. The output should look like this (you do not have to use the same variable names as I did):

```
   Obs patid lname      fname gender     bdate       vdate wgt sbp dbp

    1   001  Smith      John    M     01/25/1942 02/15/2015 189 145  82
    2   001  Smith      John    M     01/25/1942 06/25/2015 195 415  78
    3   001  Smith      John    M     01/25/1942 12/01/2015 298 138  75
    4   002  Buonamici  Sarah   F     03/17/1974 07/15/1915 142  96  62
```

2. Display all records in the combined dataset in which a visit date is earlier than the patient's birthdate. Submit your SAS code and output.

3. Use a SAS procedure to create the following table. Submit your SAS code and output.

```
--------------------------------------------------------
|                |           |  Systolic | Diastolic  |
|                |           |   blood   |   blood    |
|                |  Weight   |  pressure |  pressure  |
|                |-----------+-----------+------------|
|                |   Mean    |    Mean   |    Mean    |
|----------------+-----------+-----------+------------|
|patid           |           |           |            |
|----------------|           |           |            |
|001             |     227.33|     232.67|       78.33|
|----------------+-----------+-----------+------------|
|002             |     142.00|      96.00|       62.00|
--------------------------------------------------------
```

3 Relational database concepts

An agricultural engineering consulting firm assigns its engineers to client projects. Each engineer may work on more than one project, and usually more than one engineer works on each project. Each project has one project leader, who is one of the engineers. Each project is for only one client, but a single client may have more than one project.

The CEO of the firm wants to store the following information about his firm's engineers and projects. Place these data into a relational database structure that is in third normal form. Indicate all primary and foreign keys.

```
Project name
Project start date
Project completion date
Project leader
Project client
Project type
All engineers working on the project

Engineer name
Engineer date of hire
Engineer office phone number

Client name
Client address
Client phone number
```