Stats 20 Fall 2014

Midterm

- Solve two problems in 50 minutes. Upload your R script on CCLE by 10am.
- Not allowed to chat or email during the test.

1. It is well known that a sampling distribution of sample means is normal when we have a large sample size, even though the population is not really normal. (Central Limit Theorem) The sampling distribution is summarized:
$\bar{X} \sim N(\mu,\ \sigma/\sqrt{n})$.
Now we want to verify this theorem with R simulation. Suppose that a population is given with a command below.

> population <- rgamma(10000, 1, 1) # generate 10000 random numbers from gamma distribution

To know a distribution of the population, check a histogram and calculate a mean(μ) and standard deviation(σ) of the population.

> hist(population) # The distribution of population is right skewed.
> mean(population) # Population mean is around 1.
> sd(population) # Population standard deviation is also around 1.

a) First, we sample 100 numbers from the 'population' without replacement. What is a mean of this sample of 100 numbers (sample mean)?

b) We just got one sample mean. Now we want to have 1000 sample means by repeating a). (Each time, you will sample 100 numbers from the population and calculate a sample mean.) **Write a function *clt* with arguments of 'population', 'sample size' and 'number of repetition' that returns a list of a mean and standard deviation of sample means, sample size, and number of repetition. Try *clt(population, 100, 1000)* and you may get a similar(not same numbers) result as below.**

c) Check from the result in b) if the mean of the 1000 sample means is close to population mean and standard deviation of the sample means is close to population standard deviation over square root of 100. (That's what CLT is about.)

```
> clt(population, 100, 1000)
$mean
[1] 1.004718

$sd
[1] 0.09728716

$nsample
[1] 100

$nrepeat
[1] 1000
```

2. Explore a built-in data 'CO2'. (*?CO2* gives you a description of the data.)

a) We want to see the averages of CO2 uptake rates(uptake) for different Plants, Type and Treatment. Show how to find them. (You may get the same result as below)

	Plant	Type	Treatment	mean_uptake
1	Qn1	Quebec	nonchilled	33.22857
2	Qn2	Quebec	nonchilled	35.15714
3	Qn3	Quebec	nonchilled	37.61429
4	Qc1	Quebec	chilled	29.97143
5	Qc3	Quebec	chilled	32.58571
6	Qc2	Quebec	chilled	32.70000
7	Mn3	Mississippi	nonchilled	24.11429
8	Mn2	Mississippi	nonchilled	27.34286
9	Mn1	Mississippi	nonchilled	26.40000
10	Mc2	Mississippi	chilled	12.14286
11	Mc3	Mississippi	chilled	17.30000
12	Mc1	Mississippi	chilled	18.00000

b) We want to export the summary of the data in 1) into a file 'summary.txt' as below. Write a command to get this text file.

```
The average CO2 uptakes for the plant Qn1 in Quebec with nonchilled is 33.228571.
The average CO2 uptakes for the plant Qn2 in Quebec with nonchilled is 35.157143.
The average CO2 uptakes for the plant Qn3 in Quebec with nonchilled is 37.614286.
The average CO2 uptakes for the plant Qc1 in Quebec with chilled is 29.971429.
The average CO2 uptakes for the plant Qc3 in Quebec with chilled is 32.585714.
The average CO2 uptakes for the plant Qc2 in Quebec with chilled is 32.700000.
The average CO2 uptakes for the plant Mn3 in Mississippi with nonchilled is 24.114286
The average CO2 uptakes for the plant Mn2 in Mississippi with nonchilled is 27.342857
The average CO2 uptakes for the plant Mn1 in Mississippi with nonchilled is 26.400000
The average CO2 uptakes for the plant Mc2 in Mississippi with chilled is 12.142857.
The average CO2 uptakes for the plant Mc3 in Mississippi with chilled is 17.300000.
The average CO2 uptakes for the plant Mc1 in Mississippi with chilled is 18.000000.
```