Advanced Data Analysis 1, Stat 427/527 Homework 3 Name: _____
Fall 2014 Prof. Erik B. Erhardt

Part I. (130 points) Do all calculations in R. All R code for the assignment should be included with the part of the problem it addresses (for code and output use a fixed-width font, such as Courier). Code is used to calculate result. Text is used to report and interpret results. Do not report or interpret results in the code. Also:

1. Clearly define population parameters in each problem. That is, give a verbal description of what the population mean is in the context of the problem.
2. Clearly specify hypotheses when appropriate (not every problem involves a test of hypothesis).
3. Write a coherent conclusion based on each CI or test.

(40$^{\text{pts}}$) **1. Cloud seeding:** Return to the cloud seeding problem of Homework 2. This really is a two-sample problem, although we analyzed it in HW 2 as two one-sample problems.

unseeded	seeded
1202.6	2745.6
830.1	1697.8
372.4	1656
345.5	978
321.2	703.4
244.3	489.1
163	430
147.8	334.1
95	302.8
87	274.7
81.2	274.7
68.5	255
47.3	242.5
41.1	200.7
36.6	198.6
29	129.6
28.6	119
26.3	118.3
26.1	115.3
24.4	92.4
21.7	40.6
17.3	32.7
11.5	31.4
4.9	17.5
4.9	7.7
1	4.1

Read the data from the website with:

```
d1 <- read.csv("http://statacumen.com/teach/ADA1/ADA1_HW_02_F14-1.csv")
```

(a) (10 pts) Carefully check the assumption of normality of the data on the original scale by describing the shape of the data distribution and the sampling distribution of the mean (using the bootstrap). You need to do the seeded and unseeded days separately.

(b) (10 pts) Repeat the previous question for the log-transformed data.

(c) (20 pts) Compare the groups using two-sample t-procedures. Choose the most appropriate scale (natural or log units) in which to perform this analysis.

(30$^{\text{pts}}$) **2. Acid:** Use the Acid data (from HW 2).

Acid1	A1-2	A1-3	A1-4	Acid2
0.123	0.110	0.112	0.126	0.109
0.109	0.110	0.123	0.110	0.111
0.110	0.110	0.110	0.109	0.110
0.109	0.090	0.109	0.114	0.110
0.112	0.109	0.110	0.110	0.105
0.109	0.111	0.109	0.110	0.110
0.110	0.098	0.110	0.110	0.111
0.110	0.109	0.109	0.110	0.110
0.110	0.109	0.110	0.110	0.110
0.112	0.109	0.110	0.111	0.111
0.110	0.109	0.111	0.107	0.109
0.101	0.111	0.111	0.110	0.111
0.110	0.109	0.109	0.107	0.109
0.110	0.108	0.107	*	0.112
0.110	0.110	0.120	*	0.109
0.110	0.112	0.133	*	0.109
0.106	0.111	0.107	*	0.111
0.115	0.110	0.103	*	0.110
0.111	0.111	0.111	*	0.112
0.110	0.111	0.110	*	0.112
0.107	0.107	0.122	*	0.109
0.111	0.111	0.109	*	0.110
0.110	0.112	0.108	*	0.110
0.113	0.105	0.109	*	0.109
0.109	0.109	0.109	*	0.113
0.108	0.109	0.114	*	0.108

```
0.109  0.110  0.107  *       0.105
0.111  0.110  0.104  *       0.110
0.104  0.109  0.110  *       0.109
0.114  0.110  0.114  *       0.109
0.110  0.104  0.107  *       0.110
0.110  0.111  0.101  *       0.110
0.110  0.110  0.111  *       0.110
0.113  0.111  0.109  *       0.104
0.114  0.109  0.110  *       0.109
0.110  0.110  0.111  *       0.110
0.110  0.111  0.110  *       0.111
```

Read the data from the website with:

```
d2 <- read.csv("http://statacumen.com/teach/ADA1/ADA1_HW_02_F14-3.csv")
```

(a) (10 pts) Check the normality assumption for both experiments as in problem 1 above.

(b) (20 pts) Formally compare the experiments using two-sample t-procedures.

(60 pts) **3. cAMP:** Cyclic adenosine monophosphate (cAMP) is a substance that can mediate cellular response to hormones. In a study of maturation of egg cells in the frog Xenopts laevis, oocytes from each of four females were divided into two batches: one batch was exposed to progesterone and the other was not. After two minutes, each batch was assayed for its cAMP content, with the results given in the table below.

	cAMP (pmol/oocyte)	
Frog	Control	Progesterone
1	6.01	5.23
2	2.28	1.21
3	1.51	1.40
4	2.12	1.38

Read the data from the website with:

```
d3 <- read.csv("http://statacumen.com/teach/ADA1/ADA1_HW_03_F14-3.csv")
```

(a) (10 pts) Make a histogram and box plot of the differences between the cAMP levels for the control and progesterone samples.

(b) (20 pts) Test at the 10% level whether there is any difference in the population mean cAMP levels for batches of oocytes that are untreated versus those treated with progesterone.

(c) (10 pts) Compute and interpret a 90% CI for the difference in population mean cAMP levels for batches of oocytes that are untreated versus those treated with progesterone.

(d) (10 pts) Discuss any statistical assumptions that you have made in carrying out the analysis, and whether the assumptions seem reasonable.

(e) (10 pts) Write a short summary to the problem.