Advanced Data Analysis 1, Stat 427/527 Homework 6 Name: _____
Fall 2014 Prof. Erik B. Erhardt

Part I. (175 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.

(25pts) **1. Parallax:** The following determinations of the parallax of the sun (the angle spanned by the earth's radius as if it were viewed and measured from the sun's surface) were made in 1761 by noted astronomer James Short. The units are in seconds of a degree (1/360 degree).

```
 8.50   8.06   8.65   9.71   8.80   7.99   8.50   8.43   8.35
 8.50   8.40   8.58   7.33   8.44   8.71   8.28   8.82   8.34
 8.64   8.14   8.31   9.87   9.02   9.64   9.27   7.68   8.36
 8.86  10.57   8.34   9.06  10.34   8.58   5.76   9.11   8.55
 9.25   8.07   7.80   8.44   8.66   9.54   9.09   8.36   7.71
 8.23   8.34   9.07   8.50   9.71   8.30   8.50   8.60
```

```
parallax <- read.csv("http://statacumen.com/teach/ADA1/ADA1_HW_06_F14-1.csv")
angle <- parallax$angle;
```

With a careful determination of the radius of the earth and a good average value of the parallax, the average distance of the earth to the sun can be obtained. The currently accepted value of the parallax is 8.798.
Within this framework, define

$$\mu = \text{the population mean of all potential measurements of the parallax using Short's device.}$$

We are interested in whether μ could be the currently accepted value of the parallax of the sun, that is, we wish to test whether $\mu = 8.798$.

(a) (5 pts) Describe the distribution of determinations of the parallax. Be complete.

(b) (10 pts) Perform the standard t-test on these data, at the 5% level, and construct a 95% CI for μ. Interpret the results, given the question of interest.

(c) (10 pts) Repeat the analysis using a suitable non-parametric method, and contrast the results with part (b). Which analysis seems most reasonable, and what are your conclusions based on that analysis, given the question of interest?

(50pts) **2. Guinea pigs:** The data below are the survival times in hours of 72 guinea pigs after they were injected with a given dose of tubercule bacilli in a medical experiment. The data are from the article "Acquisition of resistance of guinea pigs injected with different doses of virulent tubercule bacilli," by T. Bjerkedal in the American Journal of Hygiene, (1960), pp. 130–148.

```
 43   45   53   56   56   57   58   66   67   73   74
 79   80   80   81   81   81   82   83   83   84   88
 89   91   91   92   92   97   99   99  100  100  101
102  102  102  103  104  107  108  109  113  114  118
121  123  126  128  137  138  139  144  145  147  156
162  174  178  179  184  191  198  211  214  243  249
329  380  403  511  522  598
```

```
guinea <- read.csv("http://statacumen.com/teach/ADA1/ADA1_HW_06_F14-2.csv")
hours <- guinea$hours;
```

(a) (10 pts) Obtain a 95% t-CI for the mean survival time.

(b) (10 pts) Repeat part (a) using a suitable nonparametric method.

(c) (10 pts) Take the log of survival time and find a 95% t-CI for mean log survival time.

(d) (10 pts) Repeat part (c) using a suitable nonparametric method.

(e) (10 pts) Compare your 4 CIs, and contrast the nonparametric with the t-CIs. If they differ much, explain why they differ. Which analysis appears most appropriate? Explain.

(50 pts) **3. Humerus sparrows:** In an 1898 Biology lecture at Woods Hole, Massachusetts, Hermon Bumpus reminded the audience that the process of natural selection for evolutionary change was an unproved theory. As evidence in support of natural selection, he presented measurements on house sparrows brought to his Brown University laboratory after an uncommonly severe winter storm. Some of the birds had died, and some had survived. Bumpus asked whether those that perished did so because they lacked physical characteristics enabling them to withstand the intensity of that particular instance of selective elimination.

The data we will look at are the humerus (arm bone) lengths (inch/1000) for the 24 adult male sparrows that perished and the 35 adult males that survived. You will see that the data are in two columns. The first column contains the humerus lengths for the 59 birds. The second column identifies whether the birds perished (0) or survived (1).

humerus	survived	humerus	survived
659	0	687	1
689	0	703	1
703	0	709	1
702	0	715	1
709	0	728	1
713	0	721	1
720	0	729	1
729	0	723	1
726	0	728	1
726	0	723	1
720	0	726	1
737	0	728	1
739	0	736	1
731	0	733	1
738	0	730	1
736	0	733	1
738	0	730	1
744	0	739	1
745	0	735	1
743	0	741	1
754	0	741	1
752	0	749	1
752	0	741	1
765	0	743	1
		741	1
		752	1
		752	1
		751	1
		756	1
		755	1
		766	1
		767	1
		769	1
		770	1
		780	1

```
sparrows <- read.csv("http://statacumen.com/teach/ADA1/ADA1_HW_06_F14-3.csv")
sparrows$survived <- factor(sparrows$survived)
```

(a) (10 pts) Make appropriate graphical displays to compare the humerus lengths in the two samples

(b) (10 pts) Test at the 5% level whether there is any difference in the population mean humerus lengths for those that perished and those that survived. Use both the t-test and an appropriate nonparametric procedure.

(c) (10 pts) Compute and interpret a 95% CI for the difference in population mean humerus lengths for those that perished and those that survived. Repeat for an appropriate nonparametric procedure.

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

(e) (10 pts) Write a short summary for the problem. What analysis seems most appropriate?

(50 pts) **4. Protoporphin levels among alcoholics:** Protoporphrin levels were determined for three groups of people — a control group of normal workers, a group of alcoholics with sideroblasts in their bone marrow, and a group of alcoholics without sideroblasts. The given data appeared in the paper "Erythocyte Coproporphyrin and Protoporphrin in Ethanol-Induced Sideroblastic Erythoporiesis" (Blood, 1974, p. 291–295).

Analyze the data, assuming you are interested in comparing the typical protoporphrin level across groups. Quantify any differences you find. Make sure to clearly define all population parameters, and assess the assumptions underlying your chosen method of analysis. Have a well-organized write-up.

Normal	Alc_w_sb	Alc_wo_sb

22	78	37
27	172	28
47	286	38
30	82	45
38	453	47
78	513	29
28	174	34
58	915	20
72	84	68
56	153	12
30	780	37
39	NA	8
53	NA	76
50	NA	148
36	NA	11

(At `statacumen.com/teach/ADA1/ADA1_HW_06_F14-4.txt`, it's up to you to read this dataset into R.)