Advanced Data Analysis 1, Stat 427/527   Name:

1. 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
```

The data is located at

http://www.math.unm.edu/~luyan/ADA118/guinea.csv

   (a) (10 pts) Obtain a 95% t-CI for the mean survival time. Check histogram, boxplot, QQplot for normality of the population distribution.
   (b) (10 pts) Repeat part (a) using a suitable nonparametric method.
   (c) (10 pts) Take the log of the survival time and construct a 95% t-CI for the mean log survival time.
   (d) (10 pts) Repeat part (c) using a suitable nonparametric method.
   (e) (10 pts) Compare the 4 CIs. If they differ much, explain why they differ. Which analysis appears most appropriate? Explain.

2. 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://www.math.unm.edu/~luyan/ADA118/sparrow.csv'')
sparrows$survived<-factor(sparrows$survived)
```

(a) Make appropriate graphical displays such as boxplot, histogram to compare the humerus lengths in the two samples

(b) 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) 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) Discuss any statistical assumptions that you have made in carrying out the t test and nonparametric procedure, and whether the assumptions seem reasonable.

(e) What analysis seems most appropriate?

3. 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. 291295).

```
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
```

The data is located at

`http://www.math.unm.edu/~luyan/ADA118/blood.txt`

(a) (10 pts)Analyze the data using both a Kruskal-Wallis test and t-test, assuming you are interested in comparing the typical protoporphrin level across groups.

(b) (10 pts) Make sure to clearly identify all population parameters, and assess the assumptions underlying both methods of analysis.

(c) (10 pts) Is it more appropriate to do the t-test on the original data or using a log transformation?

(d) (10 pts) Interpret the results of the KW and t-tests.

(e) (10 pts) In addition, do Bonferroni comparisons for pairwise differences using both Kruskal-Wallis and the t-test.