STATISTICS 641 - ASSIGNMENT 4

DUE DATE: NOON (CDT), FRIDAY, February 20, 2015

Name _____

Email Address _____

Please TYPE your name and email address. Often we have difficulty in reading the handwritten names and email addresses. Make this cover sheet the first page of your Solutions.

STATISTICS 641 - ASSIGNMENT #4 - NOON (CDT) Tuesday - 02/20/2015

• Read Handouts 6 & 7 and Chapters 2 & 4 in the Textbook.

• Submit for grading the following problems:

I. (45 points) A researcher is studying the relative brain weights (brain weight divided by body weight) for 51 species of mammal whose average litter size is less than 2 and for 45 species of mamma whose average litter size is greater than or equal to 2. The researcher was interested in determining what evidence that brain sizes tend to be different for the two groups. (Data from *The Statistical Sleuth* by Fred Ramsey and Daniel Schafer).

RELATIVE BRAIN WEIGHTS - SMALL LITTER SIZE

0.42	0.86	0.88	1.11	1.34	1.38	1.42	1.47	1.63
1.73	2.17	2.42	2.48	2.74	2.74	2.79	2.90	3.12
3.18	3.27	3.30	3.61	3.63	4.13	4.40	5.00	5.20
5.59	7.04	7.15	7.25	7.75	8.00	8.84	9.30	9.68
10.32	10.41	10.48	11.29	12.30	12.53	12.69	14.14	14.15
14.27	14.56	15.84	18.55	19.73	20.00			

RELATIVE BRAIN WEIGHTS - LARGE LITTER SIZE

0.94	1.26	1.44	1.49	1.63	1.80	2.00	2.00	2.56
2.58	3.24	3.39	3.53	3.77	4.36	4.41	4.60	4.67
5.39	6.25	7.02	7.89	7.97	8.00	8.28	8.83	8.91
8.96	9.92	11.36	12.15	14.40	16.00	18.61	18.75	19.05
21.00	21.41	23.27	24.71	25.00	28.75	30.23	35.45	36.35

1. For the Large Litter Size mammals, Compute a 10% trimmed mean, and compare it to the untrimmed sample mean. Does this comparison suggest any extreme values in the data?
2. The researcher suggested a Weibull distribution to model the data for the Large Litter Size mammals. Assuming that the Weibull distribution is an appropriate model for the Large Litter Size data, obtain the MLE estimates of the Weibull parameters for the Large Litter Size data.
3. Estimate the probability that a randomly selected mammal with a litter size of 5 will have a relative brain weight greater than 30.
4. Compare the MLE estimates of μ and σ based on the Weibull model to the distribution-free estimates of μ and σ for the Large Litter Size data?
5. Compare the MLE estimates of median and IQR based on the Weibull model to the distribution-free estimates of median and IQR for the Large Litter Size data?
6. Without any assumed model, estimate the mean and standard deviation of the relative brain weights for both Large and Small litter sizes.
7. Estimate the median and MAD of the relative brain weights for both Large and Small litter sizes.
8. Based on your plots from Assignment #3, which pair of estimates of the center and spread in the two data sets best represents the center and spread in the two populations of relative brain weights?
9. Using your answers from the previous three questions, suggest a relationship (if any) between litter size and relative brain weights.

II. (20 points) The following data is the monthly average of daily yields of Moody's AAA bonds for the years 1989 to 1993.

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
1989	9.62	9.64	9.80	9.79	9.57	9.10	8.93	8.96	9.01	8.92	8.89	8.96
1990	8.99	9.22	9.37	9.46	9.47	9.26	9.24	9.41	9.56	9.53	9.30	9.05
1991	9.04	8.83	8.93	8.86	8.86	9.01	9.00	8.75	8.61	8.55	8.48	8.31
1992	8.20	8.29	8.35	8.33	8.28	8.22	8.07	7.95	7.92	7.99	8.10	7.98
1993	7.91	7.71	7.58	7.46	7.43	7.33	7.17	6.85	6.66	6.67	6.93	6.90

1. Create a time series plot of the data.
2. Calculate the values of ρ_k, the autocorrelation coefficients. What conclusions can you draw?
3. Does the time series appear to be stationary? That is, do the mean and variance appear to remain constant over time.

III. (20 points) Twenty-five patients diagnosed with rare skin disease are randomly assigned to two drug treatments. The following times are either the time in days from the point of randomization to either a complete recovery or censoring (as indicated by the status variable: 0 means censored, i.e., time at which patient left study prior to a complete recovery, 1 means patient's time to recovery).

	Treatment 1												
Time	180	632	2240	195	76	70	13	1990	18	700	210	1296	23
Status	1	1	1	1	1	1	0	0	1	1	1	1	1
	Treatment 2												
Time	8	852	52	220	63	8	1976	1296	1460	63	1328	365	
Status	0	1	1	1	1	1	0	0	1	1	1	1	

1. Estimate the survival function for the two treatments.
2. Compare the mean and median time to death for the two treatments.
3. Which treatment appears to be most effective in the treatment of the skin disease?

IV. (15 points) **Select** the letter of the **BEST** answer. Justify your answer with at most 20 words.

1. An experiment involves putting specimens of steel under stress until the specimen fractures. The machine increases the stress until the specimen fractures. The maximum stress that the machine can place on a specimen is 500 psi. Out of the 35 specimens used in the experiment, 5 did not fracture at 500 psi. This type of censoring is called
   A. Random censoring
   B. Type I censoring
   C. Type II censoring
   D. Left censoring
   E. Right censoring

2. A veterinarian designed a study to determine the age at which Labrador retrievers learned how to swim. There was three groups of puppies:

   Group I: Puppies who knew how to swim prior to the beginning of the study;

   Group II: Puppies who learned how to swim during the study;

   Group III: Puppies who had not yet learned how to swim at the conclusion of the study.

   The age at which each puppy learned how to swim was recorded. The values recorded for the Group I puppies are

   A. Type I censored
   B. Type II censored
   C. Random censored
   D. Left censored
   E. Uncensored

3. Refer to Problem 2. The values recorded for the Group II puppies are

   A. Type I censored
   B. Type II censored
   C. Random censored
   D. Left censored
   E. Uncensored

4. Refer to Problem 2. The values recorded for the Group III puppies are

   A. Type I censored
   B. Type II censored
   C. Random censored
   D. Left censored
   E. Uncensored

V. Prove the following statement (5 bonus points)

If Y has a symmetric distribution with $\mu < \infty$ and median $\tilde{\mu}$,

then, the median of $W = |Y - \tilde{\mu}|$, equals the SIQR of Y.