# STATISTICS 641 - ASSIGNMENT 4

## DUE DATE: NOON (CDT), WEDNESDAY, OCTOBER 6, 2021

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) Wednesday - 10/6/2021**

- Read: Handouts 6 and 7
- Supplemental Reading: Chapter 1 & Sections 4.6, 6.1 in Devore book and *Applied Survival Analysis Using R*
- Submit for grading the following problems:

P1. ( 50 points) A researcher is studying the relative brain weights (brain weight divided by body weight) for 51 species of mammals whose litter size is 1 and for 44 species of mammals 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 | |

1. For the Small Litter Size mammals, answer the following questions: The data is given in the file: Brain Weight Data.txt in Canvas
   a. Compute a 10% trimmed mean, and compare it to the untrimmed sample mean. Does this comparison suggest any extreme values in the data?
   b. The researcher suggested a Weibull distribution to model the data for the Small Litter Size mammals. Assuming that the Weibull distribution is an appropriate model for the Small Litter Size data, obtain the MLE estimates of the Weibull parameters for the Small Litter Size data.
   c. Estimate the probability that a randomly selected mammal with a litter size of 1 will have a relative brain weight greater than 15, first using the Weibull model and secondly using a distribution-free estimate.
   d. Compare the MLE estimates of $\mu$ and $\sigma$ based on the Weibull model to the distribution-free estimates of $\mu$ and $\sigma$ for the Small Litter Size data.
   e. Compare the MLE estimates of median and IQR based on the Weibull model to the distribution-free estimates of median and IQR for the Small Litter Size data.
2. Without any assumed model, estimate the mean and standard deviation of the relative brain weights for both Large and Small litter sizes.
3. Estimate the median and MAD of the relative brain weights for both Large and Small litter sizes.
4. 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?
5. Using your answers from the previous three questions, suggest a relationship (if any) between litter size and relative brain weights.

P2. ( 30 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. Compute the mean and median time to recovery for the two treatments using the estimated survival function.
3. Which treatment appears to be most effective in the treatment of the skin disease?
4. Estimate the mean and median time to recovery ignoring the censoring and compare these values to values obtained in part 2.

P3. (20 points) **Select** the letter of the **BEST** answer.

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. Right censoring
   B. Type I censoring
   C. Type II censoring
   D. Random censoring
   E. Left censoring

2. An entomologist is interested in the ability of ticks to conserve water in very dry condition, relative humidity less than 10%. She randomly selects 100 Lone Star ticks for a large collection of Lone Star ticks and places them in a water-free container in which the temperature is maintained at $30°C$ with a relative humidity of 10%. The amount of water in the ticks will gradually decline over time. The amount of water retained by the ticks is measured after 90 days. Twelve of the 100 ticks did not survive until the end of the study but their water contents were recorded at the time of their death.

   We would describe the data from this study as being

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

3. A chemist employed at a large cosmetic firm designs a study to assess the toxicity of a new skin conditioner. He simultaneously feeds 100 mice a large volume of the conditioner. The time to death is recorded for the mice. The study was terminated after 30 days at which time twelve of the mice were still alive. The data from this type of study is best described as having

   A. Right censoring
   B. Type I censoring
   C. Type II censoring
   D. Random censoring
   E. Left censoring

4. The product engineer for an automobile safety testing agency is evaluating the likelihood of a fire in the batteries of electric automobiles. She randomly selects 100 electric vehicles for testing and the cars were driven by the employees of the agency. The study was terminated when the 20th vehicle had a fire in its battery. The engineer recorded the number of miles each vehicle was driven until either the vehicle's battery caught on fire or until the study was terminated. What type of censoring took place, if any?

   A. Right censoring
   B. Type I censoring
   C. Type II censoring
   D. Random censoring
   E. There is no censoring because a mileage was recorded for each vehicle.

5. An engineer for an automotive manufacturer is studying the occurrence of a defective in the braking system for a newly designed braking system. She randomly selects 100 automobiles for study and plans to record the distance traveled prior to a failure in the braking system. However, she needs to conclude the study 12 months after its inception. For each of the 100 automobiles she recorded the mileage at which a failure occurred in the braking system or the mileage driven during the 12 month study for those automobiles that did not have a failure. We would describe the data from this type of study as having

   A. Right censoring
   B. Type I censoring
   C. Type II censoring
   D. Random censoring
   E. Left censoring

**Bonus Problems for 10 points (attempt problems only if you have extra time).**

A. Bonus Problem 1 (5 points) Let a random variable $Y$ have a continuous strictly increasing cdf $F$ with pdf $f$.

Let $\mu_{(\alpha)}$ be the $\alpha-$trimmed mean of $Y$, that is,

$$\mu_{(\alpha)} = \frac{1}{1-2\alpha} \int_{Q(\alpha)}^{Q(1-\alpha)} y f(y) dy.$$

Prove that

$$\lim_{\alpha \to .5} \mu_{(\alpha)} = \tilde{\mu}, \qquad \text{where} \quad \tilde{\mu} = Q(.5), \quad \text{the median of the distribution of } \; Y$$

Hint 1: Use l'Hôpital's rule in your proof and

the fact that $\frac{d}{dx} \int_{g(x)}^{h(x)} t f(t) dt = h(x) f(h(x)) h^{'}(x) - g(x) f(g(x)) g^{'}(x)$

Hint 2: Also use the fact that $F(Q(u)) = u \Rightarrow \frac{d}{du} F(Q(u)) = \frac{d}{du} u = f(Q(u)) Q'(u) = 1$

B. Bonus Problem 2 (5 points)

Let $T_1, T_2, \ldots, T_{20}$ be the miles to failure of 20 wheel bearings in 20 newly manufacturered trucks. The values were obtained in an accelerated life testing study and are as follows in units of 1000 miles of use:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 37.52 | 30.33 | 42.82 | 31.14 | 31.49 | 38.04 | 38.31 | 36.15 | 34.08 | 30.76 |
| 45.14 | 44.81 | 33.30 | 45.73 | 30.00 | 33.97 | 33.65 | 30.60 | 43.07 | 31.26 |

The 20 values are modeled as a shifted exponential distribution with p.d.f. given as follows:

$$g(t; \beta, \theta) = \beta e^{-\beta(t-\theta)} \quad \text{for} \quad t \geq \theta$$

i. Write the likelihood function of $\beta$ and $\theta$

ii. Determine the MLE of $\beta$ and $\theta$ assuming both parameters are unknown

iii. Estimate the probability that the miles to failure for a randomly selected truck is greater than 40,000 miles. Note that $G(t; \beta, \theta) = 1 - e^{\beta(t-\theta)}$.