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 μ and σ based on the Weibull model to the distribution-free estimates of μ and σ 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,

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

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

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 β and θ

ii. Determine the MLE of β and θ 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 $G(t; \beta, \theta) = 1 - e^{\beta(t-\theta)}$.

Stat 641 Fall 2021

Solutions for Assignment 4

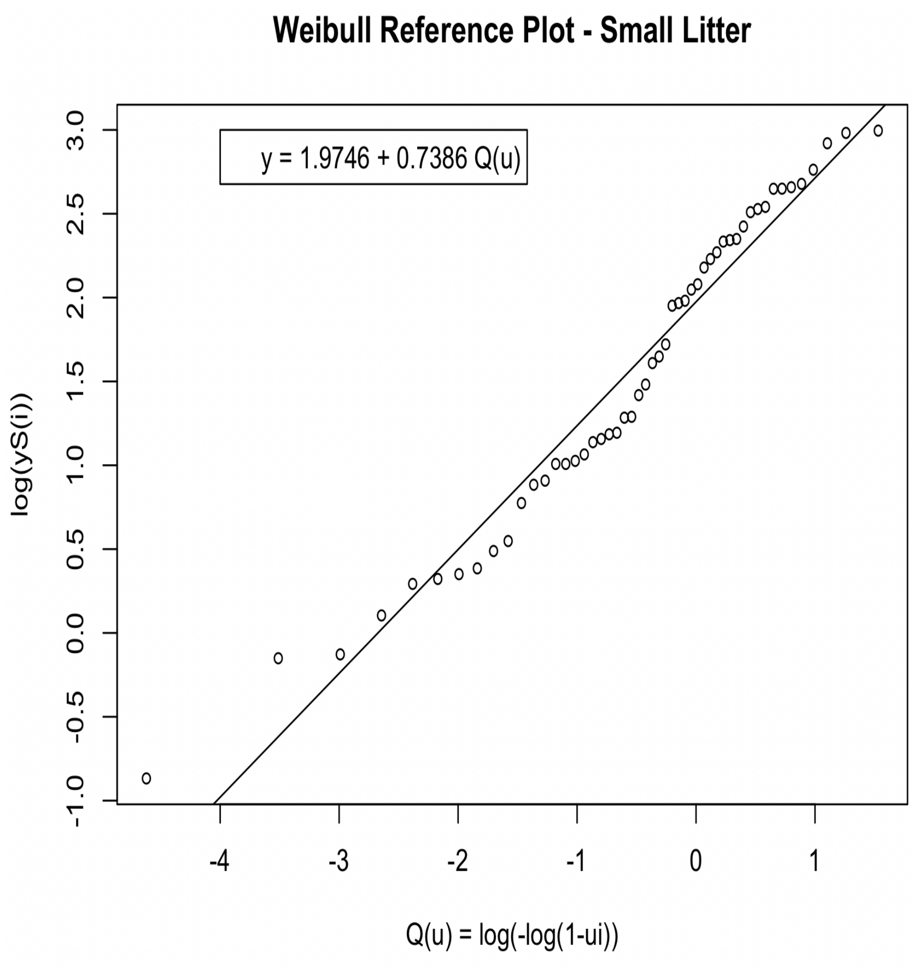
P1. (50 points)

1. See R code at the end of this document. For the Small Litter Size we obtain:

   a. A 10% trimmed mean would involve averaging the middle
      $K(.10) = 51 - [(51)(.1)] - [(51)(.1) + 1] + 1 = 51 - [5.1] - [6.1] + 1 = 41$ values in the data set yielding:
      $\hat{\mu}_{(.1)} = \frac{1}{41}\sum_{i=6}^{46} Y_{(i)} = \frac{1}{41}(257.89) = 6.29 = mean(yS, trim = .1)$ whereas the untrimmed mean is
      $\hat{\mu} = \frac{1}{51}\sum_{i=1}^{51} Y_{(i)} = \frac{1}{51}(351.18) = 6.89$

      The untrimmed mean is somewhat larger than the 10% trimmed mean which would indicate that there a few large outliers in the data. In fact, examining the sorted data, we have three relatively large data values in the data set: 18.55, 19.73, 20.00

   b. A Weibull reference distribution plot is displayed:



The plot indicates a reasonably good fit of the Small Litter data to a Weibull distribution. The graphical estimates are

$$\hat{\gamma} = 1/.7386 = 1.3539 \qquad \hat{\alpha} = e^{1.9746} = 7.2037$$

The MLE of the Weibull parameters from R are

$$\hat{\gamma} = \text{Weibull Shape} = 1.265583 \qquad \hat{\alpha} = \text{Weibull Scale} = 7.4268$$

A fairly good match to the graphical estimates.

c. Using the MLE estimates: $P[Y_L > 15] = 1 - F(15) \approx e^{-(15/7.4268)^{1.265583}} = .0877$

Using the Graphical estimates: $P[Y_L > 15] = 1 - F(15) \approx e^{-(30/7.2037)^{1.3539}} = .0672$ about 23% less than the MLE

The distribution-free estimate would be $P[Y > 15] = 1 - \hat{F}(15) = 1 - 47/51 = .0784$ about 11% less than the MLE

d. The distribution-free estimates are

$\hat{\mu}$ = sample mean $= \bar{Y}_S = 6.8859$ $\quad\hat{\sigma}$ = sample stand. dev. $= S_{Y_S} = 5.46$

Using the formulas on page 6 in HO 6, we have for the Weibull distribution, with MLE's from R:

$$\widehat{\mu} = \widehat{\alpha}\Gamma\left(1 + \frac{1}{\widehat{\gamma}}\right) = (7.4268)\Gamma\left(1 + \frac{1}{1.265583}\right) = 6.8981$$

$$\widehat{\sigma} = \widehat{\alpha}\sqrt{\Gamma\left(1 + \frac{2}{\widehat{\gamma}}\right) - \Gamma^2\left(1 + \frac{1}{\widehat{\gamma}}\right)} = 7.4268\sqrt{\Gamma\left(1 + \frac{2}{1.265583}\right) - \Gamma^2\left(1 + \frac{1}{1.265583}\right)} = 5.4882$$

The MLE estimates of μ and σ are very close to the distribution-free estimates thus lending evidence that the Weibull model is the correct model for this data.

e. The distribution-free estimates are

$\hat{\tilde{\mu}} = \hat{Q}(.5)$ = sample median $= Y_{(26)} = 5.00$ $\quad$ Using R-function, quantile(yS,.5,type=5) = 5.00

$\widehat{IQR}$ = sample IQR $= \hat{Q}(.75) - \hat{Q}(.25) = Y_{(.75n+.5)} - Y_{(.25n+.5)} = Y_{(38.75)} - Y_{(13.25)} \Rightarrow$

$\hat{Q}(.75) - \hat{Q}(.25) = (.25 * Y_{(38)} + .75 * Y_{(39)}) - (.75 * Y_{(13)} + .25 * Y_{(14)}) = 10.4625 - 2.545 = 7.9175$

Using R-function, $quantile(yL, .75, type = 5) - quantile(yL, .25, type = 5) = 10.4625 - 2.545 = 7.9175$

Using the formula for the quantile function from a Weibull distribution:

$Q(u) = \alpha(-log(1-u))^{1/\gamma}$ along with MLE from R for α and γ we have

$\widehat{\tilde{\mu}} = \hat{Q}(.5) = \hat{\alpha}(-log(1-.5))^{1/\hat{\gamma}} = 7.42681(-log(1-.5))^{1/1.265583} = 5.56$

$\widehat{IQR} = \hat{Q}(.75) - \hat{Q}(.25) = \hat{\alpha}(-log(1-.75))^{1/\hat{\gamma}} - \hat{\alpha}(-log(1-.25))^{1/\hat{\gamma}} = 6.8387$

Equivalently, using the R quantile function for the Weibull distribution, we have

$\widehat{\tilde{\mu}} = qweibull(.5, 1.265583, 7.42681) = 5.56$

$\widehat{IQR} = qweibull(.75, 1.265583, 7.42681) - qweibull(.25, 1.265583, 7.42681) = 6.8387$

The MLE estimate of the median based on the Weibull model is close to the distribution-free estimate (5.56 to 5.00) but there is substantial difference between the two estimates of the IQR (6.8387 to 7.9175). This may be due to the IQR reflecting only the fit of the data in the middle of the distribution.

2. For Large Litter Data: $\hat{\mu}_L = 10.39$ $\hat{\sigma}_L = 9.15$

   For Small Litter Data: $\hat{\mu}_S = 6.89$ $\hat{\sigma}_S = 5.46$

3. For Large Litter Data: $\hat{\tilde{\mu}}_L = 7.93$ $M\hat{A}D_L = 7.95$

   For Small Litter Data: $\hat{\tilde{\mu}}_S = 5.00$ $M\hat{A}D_S = 5.23$

4. For the Small Litter Data, the pdf appeared to be just slightly right skewed so the mean should be only slightly larger than the median (6.89 vs 5.00) and the standard deviation somewhat larger than MAD (5.46 vs 5.23). The larger than expected difference in the Mean and Median was very surprising considering that S and MAD were so close in value.

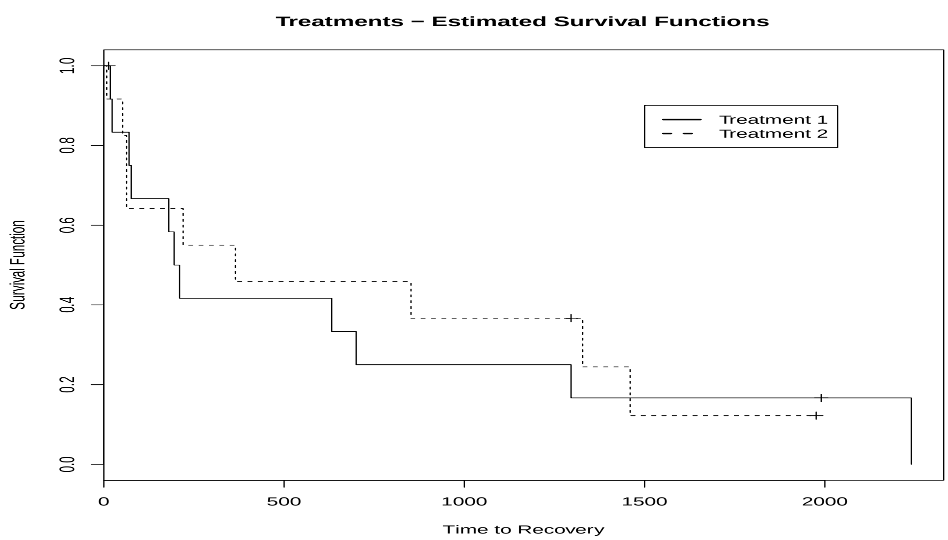
   For the Large Litter Data, the pdf appeared to be just more right skewed so the mean should be larger than the median (10.39 vs 7.93) and the standard deviation somewhat larger than MAD (9.15 vs 7.95). I was somewhat surprised that there was not a larger difference between S and MAD considering the 4 or 5 rather large values in the Large Litter data set.

Based on the right skewness of the estimated pdf for the Large Litter data and the goal of the study was to compare the Small to the Large Litter relative brain weights, I would select (Median, MAD) to represent the location and scale in the two data sets.

5. From the given data, it would appear that larger relative brain weights are associated with Larger Litter sizes. It would be much more informative to have the actual litter sizes associated with each relative brain weight as opposed to having the groupings into just small and large litters.

P2. (30 points) Using the times to recovery (or censoring) for the 25 patients we obtain:

1. The estimate survival functions for the two Treatments are given in the following plot:



2. From the R output we have

```
               G=1
time n.risk n.event survival std.err lower 95% CI upper 95% CI
  18     12       1    0.917  0.0798       0.7729        1.000
  23     11       1    0.833  0.1076       0.6470        1.000
  70     10       1    0.750  0.1250       0.5410        1.000
  76      9       1    0.667  0.1361       0.4468        0.995
 180      8       1    0.583  0.1423       0.3616        0.941
 195      7       1    0.500  0.1443       0.2840        0.880
 210      6       1    0.417  0.1423       0.2133        0.814
 632      5       1    0.333  0.1361       0.1498        0.742
 700      4       1    0.250  0.1250       0.0938        0.666
1296      3       1    0.167  0.1076       0.0470        0.591
2240      1       1    0.000     NaN           NA           NA
               G=2
time n.risk n.event survival std.err lower 95% CI upper 95% CI
   8     12       1    0.917  0.0798       0.7729        1.000
  52     10       1    0.825  0.1128       0.6311        1.000
  63      9       2    0.642  0.1441       0.4132        0.996
 220      7       1    0.550  0.1499       0.3224        0.938
 365      6       1    0.458  0.1503       0.2410        0.872
```

```
  852       5       1     0.367  0.1456       0.1684        0.798
 1328       3       1     0.244  0.1392       0.0801        0.746
 1460       2       1     0.122  0.1110       0.0206        0.724
> print(results, print.rmean=TRUE)
Call: survfit(formula = Surv(T, ST) ~ G)
    records n.max n.start events *rmean *se(rmean) median 0.95LCL 0.95UCL
G=1      13    13      13     11    657        229    202      76      NA
G=2      12    12      12      9    731        216    365      63      NA
```

Note that for G=1, the table reports the median as 202 but $\hat{S}(195) = .5$ from the output of the K-M estimator of the survival function. According to our definition of the quantile function, $\hat{Q}(u) = inf\{t : \hat{S}(t) \le 1-u\}$, the median would be 195.

The estimated mean and median are smaller for Treatment 1 (G=1) than for Treatment 2 (G=2).

3. Based on the median time to recovery, Treatment 1 would be the more effective treatment. The mean times to recovery are much larger than the median times due to a few very large values in both treatment groups. But, Treatment 1 still has a smaller mean the Treatment 2. However, as we will discuss in future handouts, when the standard errors of the estimators are taken into account, there may not be significant evidence of a difference in the two treatments.

P4. (20 points)

1. **A or B -** Because the true stress for the censored specimens are greater than or equal to $t_C = 500$ psi

2. **D -** Because the amount of water retained at 90 days would be less than the amount of water retained at death.

3. **A or B -** Because the study is terminated at a fixed time, 30 days

4. **A or C -** Because the study was terminated after a pre-selected number of fires

5. **A -** Because brake failure mileage for the censored automobiles are greater than the miles traveled at the end of the study.

V. (10 Bonus points)

- Bonus 1. (5 points)

$$\lim_{\alpha\to .5}\mu_{(\alpha)} = \frac{\lim_{\alpha\to .5}\int_{Q(\alpha)}^{Q(1-\alpha)} yf(y)dy}{\lim_{\alpha\to .5}(1-2\alpha)} = \frac{0}{0}$$

Apply $l^{'}$Hopital's Rule:

$$\begin{aligned}
\lim_{\alpha\to .5}\mu_{(\alpha)} &= \frac{\lim_{\alpha\to .5}\frac{d}{d\alpha}\int_{Q(\alpha)}^{Q(1-\alpha)} yf(y)dy}{\lim_{\alpha\to .5}\frac{d}{d\alpha}(1-2\alpha)} \\
&= \frac{\lim_{\alpha\to .5}[Q(1-\alpha)f(Q(1-\alpha))(-1)Q^{'}(1-\alpha) - Q(\alpha)f(Q(\alpha))Q^{'}(\alpha)]}{-2} \\
&= \frac{-2Q(.5)f(Q(.5))Q^{'}(.5)}{-2} = Q(.5)
\end{aligned}$$

Therefore, $Q^{'}(u) = \frac{1}{f(Q(u))}$

- Bonus 2. (5 points)

  i. The likelihood function is given by

$$L(\beta,\theta;t_1,t_2,\ldots,t_n) = \prod_{i=1}^{n} g(t_i;\ \beta,\theta) = \prod_{i=1}^{n}\beta e^{-\beta(t_i-\theta)}I(t_i\geq\theta)$$

  ii. As a function of θ, the likelihood increases as θ increase, until $\theta \geq min(t_1,t_2,\ldots,t_n)$ after which the likelihood becomes 0.

  - Therefore, the MLE for θ is $\hat{\theta} = min(t_1,t_2,\ldots,t_n)$

    The log-likelihood function is given by

$$l(\beta,\theta;t_1,t_2,\ldots,t_n) = log(L(\beta,\theta;t_1,t_2,\ldots,t_n)) = nlog(\beta) - \beta\left(\sum_{i=1}^{n}t_i - n\theta\right) \quad \text{for all} \quad t_i\geq\theta$$

    The log-likelihood evaluated at $\hat{\theta}$ is

$$l = l(\beta,\hat{\theta};t_1,t_2,\ldots,t_n) = nlog(\beta) - \beta\left(\sum_{i=1}^{n}t_i - n\hat{\theta}\right) \quad \text{for all} \quad t_i\geq\hat{\theta}$$

$$\frac{dl}{d\beta} = \frac{n}{\beta} - \sum_{i=1}^{n}(t_i-\hat{\theta}) \quad \text{and} \quad \frac{d^2l}{d\beta^2} = \frac{-n}{\beta^2} < 0$$

    Setting $\frac{dl}{d\beta}$ equal to 0 and solving for β yields

$$\hat{\beta} = \frac{n}{\sum_{i=1}^{n}(t_i-\hat{\theta})} \quad \text{which is a maximum because 2nd derivative was negative}$$

  iii. $P[T>40] = e^{-\hat{\beta}(40-\hat{\theta})} = e^{-.1637(40-30)} = 0.1946$

```
####
#### P1
####

##
## (1)
##

library(MASS)
yS = c(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)
yL =
 c( 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 )

nS <- length(yS)
nL <- length(yL)

## (a)
yS  = sort(yS)
ySt = yS[c(-(1:5),-(nS - 0:4))]
meanS = mean(yS)
trim.meanS = mean(ySt)
trimmedmean = mean(yS,trim=.10)

## (b)
i <- 1:nS
ui <- (i - 0.5) / nS
QW <- log(-log(1 - ui))
plot(QW, log(yS), main="Weibull Reference Plot - Small Litter",cex=.75,lab=c(7,11,7),
       xlab="Q(u) = log(-log(1-ui))",
       ylab="log(yS(i))")
abline(lm(log(yS) ~ QW))
legend(-4,3.0,"y = 1.9746 + 0.7386 Q(u)")

mle_weib <- fitdistr(yS,"weibull")
gamma_hat <- 1.2655827
alpha_hat <- 7.4268097

## (c)
exp(-(15 / alpha_hat) ^ gamma_hat)

## (d)
mu_hat <- alpha_hat * gamma(1 + 1 / gamma_hat)
sigma_hat <- sqrt(alpha_hat ^ 2 * (gamma(1 + 2 / gamma_hat) -
  (gamma(1 + 1 / gamma_hat)) ^ 2))

mean(yS)
sd(yS)

## (e)
med_hat <- alpha_hat * (-log(1 - 0.5)) ^ (1 / gamma_hat)
Q1_hat <- alpha_hat * (-log(1 - 0.25)) ^ (1 / gamma_hat)
Q3_hat <- alpha_hat * (-log(1 - 0.75)) ^ (1 / gamma_hat)
IQR_hat <- Q3_hat - Q1_hat

0.5 * nS + 0.5
0.25 * nS + 0.5
0.75 * nS + 0.5
med_df <- yS[26]
Q1_df <- 0.75 * yS[13] + 0.25 * yS[14]
Q3_df <- 0.25 * yS[38] + 0.75 * yS[39]
median(yS)
quantile(yS, c(0.25, 0.75), type = 5)
```

```
IQR_df <- Q3_df - Q1_df

##
## (2)
##

xbar_S <- mean(yS)
s_S <- sd(yS)
xbar_L <- mean(yL)
s_L <- sd(yL)

##
## (3)
##

med_L <- mean(yL)
iqr_L <- quantile(yL, 0.75) - quantile(yL, 0.25)
mad_L <- mad(yL)
mad_S <- mad(yS)

####
#### P2
####

library(survival)

T = c( 180, 632, 2240, 195, 76, 70, 13, 1990, 18, 700, 210, 1296, 23, 8, 852,  52, 220, 63,   8, 1976,1296,1460,63,1328,365)
ST = c(  1,   1,    1,   1,  1,  1,  0,    0,  1,   1,   1,    1,  1, 0,   1,   1,   1,  1,   1,    0,   0,   1, 1,   1,  1)
G =  c( rep(1,13),rep(2,12))

out = cbind(T,ST,G)
Surv(T, ST)

results <- survfit(Surv(T, ST) ~ G)
summary(results)
print(results, print.rmean=TRUE,rmean="individual",mark.time=True)

par(lab=c(15,20,4))
plot(results,ylab="Survival Function",xlab="Time to Recovery",mark.time=TRUE,
main="Treatments - Estimated Survival Functions",lty=1:2 )
legend(1500,.9,c("Treatment 1","Treatment 2"),lty=1:2,lwd=2)
```