```
#### STAT 641 Assignment 4 ####

# 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 thanor equal to 2.  The researcher was interested in determining
what evidence that brain sizes
# tend to be differentfor the two groups.  (Data fromThe Statistical Sleuthby
Fred Ramsey and Daniel Schafer).


BrainSize = read.csv("C:/Users/jackr/OneDrive/Desktop/Graduate School Courses/
STAT 641 - Methods of STAT I/RawData/Assign3_BrainSize.csv")
BrainSize = data.frame(SmallLitter = BrainSize$ï..Small.Litter.Size,
LargeLitter = BrainSize$Large.Litter.Size)
LLitter = na.omit(BrainSize$LargeLitter)
SLitter = BrainSize$SmallLitter

#1.)  For the Small Litter Size mammals,  answer the following questions:  The
data is given in the file:
#      BrainWeight 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?
mean(BrainSize$SmallLitter)
mean(BrainSize$SmallLitter, trim = .10)

# the trimmed mean and the untrimmed sample mean are very similar.  This would
suggest there are few (if 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,
#      obtainthe MLE estimates of the Weibull parameters for the Small Litter
Size data.

library(MASS)
mle_weibull=fitdistr(BrainSize$SmallLitter,"weibull",lower=c(0,0))

# shape:
mle_weibull$estimate[1]

# scale:
mle_weibull$estimate[2]

#(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.
```

```
# Weibull:
1 - pweibull(15, shape = mle_weibull$estimate[1], scale =
mle_weibull$estimate[2])
exp(-(15/mle_weibull$estimate[2])^mle_weibull$estimate[1])

# Distribution Free Estimate:
1-ecdf(BrainSize$SmallLitter)(15)

#(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.
lambda = mle_weibull$estimate[2]
k = mle_weibull$estimate[1]

lambda*gamma(1+1/k)
mean(BrainSize$SmallLitter)

sqrt((lambda^2)*(gamma(1+2/k) - (gamma(1+1/k)^2)))
sd(BrainSize$SmallLitter)

# the MLE estimates and the distribution free estimates are very similar

# (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.

# Distribution Free
median(BrainSize$SmallLitter)
IQR(BrainSize$SmallLitter)
quantile(BrainSize$SmallLitter)

# MLE Estimates
qweibull(.5, shape = mle_weibull$estimate[1], scale = mle_weibull$estimate[2])
qweibull(c(.25,.75), shape = mle_weibull$estimate[1], scale =
mle_weibull$estimate[2])


#2.) Without any assumed model, estimate the mean and standard deviation of
the relative brain weights forboth Large
#    and Small litter sizes.
mean(LLitter)
sd(LLitter)

mean(BrainSize$SmallLitter)
sd(BrainSize$SmallLitter)

# 3.) Estimate the median and MAD of the relative brain weights for both Large
and Small litter sizes.
absDevS = NULL
for(i in seq_along(SLitter)){
  absDevS[i] = abs(SLitter[i]-quantile(BrainSize$SmallLitter,.5))
}
quantile(absDevS,0.5)/.6745
```

```
absDevL = NULL
for(i in seq_along(LLitter)){
  absDevL[i] = abs(LLitter[i]-quantile(LLitter,.5))
}
quantile(absDevL,0.5)/.6745


# 4.) Based on your plots from Assignment #3, which pair of estimates of the
center and spread in the two
#      datasets best represents the center and spread in the two populations of
relative brain weights?

# The sample mean seems to be the best estimate of the center and the MAD
seems to be the best estimate of
# the spread because we have highly skewed data

# 5.) Using your answers from the previous three questions, suggest a
relationship (if any) between litter
#      size andrelative brain weights.

# Based on the previous three questions, I would say that there is a positive
relationship between average
# litter size and relative brain weights.


# P2. ( 30 points) Twenty-five patients diagnosed with rare skin disease are
randomly assigned to two drug
#      treat-ments. The  following  times  are  either  the  time  in  days
from  the  point  of
#      randomization  to  either  a  completerecovery or censoring (as
indicated by the status variable:
#      0 means censored, i.e., time at which patient left studyprior to a
complete recovery, 1 means patient's
#      time to recovery).



library(survival)

Time = c(180, 632, 2240, 195, 76, 70, 13, 1990, 18, 700, 210, 1296, 23)
Status = c(1,1,1,1,1,1,0,0,1,1,1,1,1)

Time2 = c(8, 852, 52, 220, 63, 8, 1976, 1296, 1460, 63, 1328, 365)
Status2 = c(0,1,1,1,1,1,0,0,1,1,1,1)

# 1.) Estimate the survival function for the two treatments.
Surv(Time, Status)
cords.surv1 <- survfit(Surv(Time, Status) ~ 1,conf.type="log-log")
summary(cords.surv1)
print(cords.surv1,print.rmean=TRUE)
```

```
Surv(Time2, Status2)
cords.surv2 <- survfit(Surv(Time2, Status2) ~ 1,conf.type="log-log")
summary(cords.surv2)
print(cords.surv2,print.rmean=TRUE)

# 2.) Compute the mean and median time to recovery for the two treatments
using the estimated survival function.

# Treatment 1: mean - 657, median - 202
# Treatment 2: mean - 731, median - 365

# 3.) Which treatment appears to be most effective in the treatment of the
skin disease?

# Treatment 1 seems to be most effective in the treatment of the skin disease
as the mean recovery
# time is lower than that of Treatment 2


# 4.) Estimate the mean and median time to recovery ignoring the censoring and
compare these values to values
#       obtained in part 2.

mean(cords.surv1$time)
median(cords.surv1$time)

mean(cords.surv2$time)
median(cords.surv2$time)

# the mean and median for treatment 1 is lower and the mean and median for
treatment 2 is higher


# P3.   (20 points)Selectthe letter of theBESTanswer.

# 1.) A
# 2.) D
# 3.) B
# 4.) C
# 5.) B
```