Homework 04

Joseph Blubaugh
jblubau1@tamu.edu
STAT 641-720

I.

1) mean = 10.968
   trimmed mean = 9.667 with $K(\alpha) = 45 - 45(.1) - (45(.1) + 1) - 1 = .1$
   This suggests that there are some extreme values in the top 10 percent that are pulling the mean up

2) $Weibull(\gamma, \alpha) = \frac{\gamma}{\alpha}(\frac{x}{\alpha})^{\gamma-1}e^{-(x/\alpha)^{\gamma}}$
   parameter estimates: $\gamma = 1.132, \alpha = 11.498$

```
    shape         scale
  1.1326673   11.4982087
( 0.1317455) ( 1.6007013)
```

3) Assuming $Weibull(\gamma = 1.1326673, \alpha = 11.4982087)$
   P(x > 30) = 0.0517

4)

```
[1] "MLE Mean Estimate: 10.987   MLE Sigma Estimate: 9.676"

[1] "Distribution Free Mean: 10.968   Distribution Free Sigma: 9.837"
```

5)

```
[1] "MLE Median Estimate: 8.383   MLE IQR Estimate: 11.545"

[1] "Distribution Free Median: 7.97   Distribution Free IQR: 15.22"
```

6)

```
[1] "Large Mean: 10.968   Large SD: 9.837"

[1] "Small Mean: 6.886   Small SD: 5.46"
```

7)

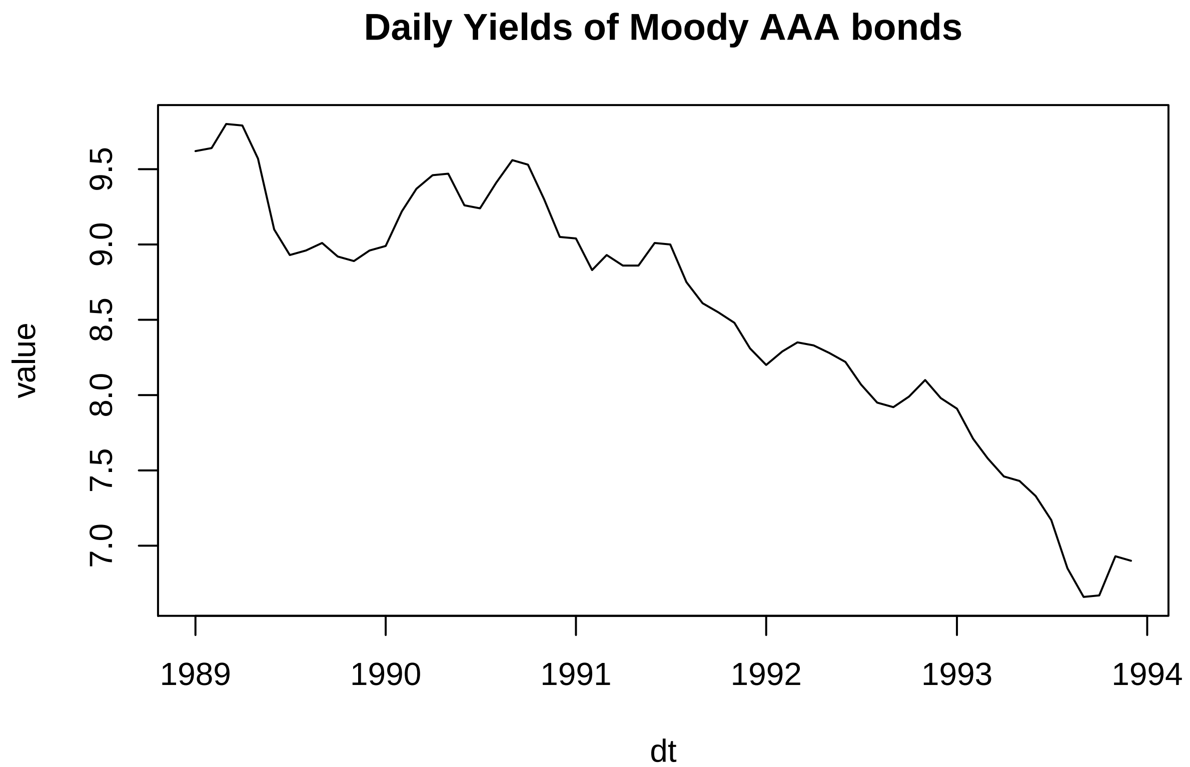
```
[1] "Large Median: 7.97   Large MAD: 8.021"

[1] "Small Mean: 5   Small MAD: 5.234"
```

8) Since both distributions are right skewed, median and MAD do the best job at describing the center and spread of the distributions.

9) Large litters on average have higher brain_wt/body_wt based on the centeroid measurements like median and mean, but large litters also have a much larger variation in brain_wt/body_wt where as the small litters brain_wt/body_wt mass is more concentrated.

II.

1)



2) You can conclude that there is a strong positive correlation between adjacent values

```
Autocorrelations of series 'ts$value', by lag

    0     1     2     3     4     5     6     7     8     9    10    11
1.000 0.939 0.864 0.775 0.690 0.618 0.568 0.524 0.482 0.438 0.404 0.378
   12    13    14    15    16    17
0.357 0.333 0.304 0.266 0.226 0.187
```

3) No, it is easy so see that the mean has been declining over the course of time.

III.

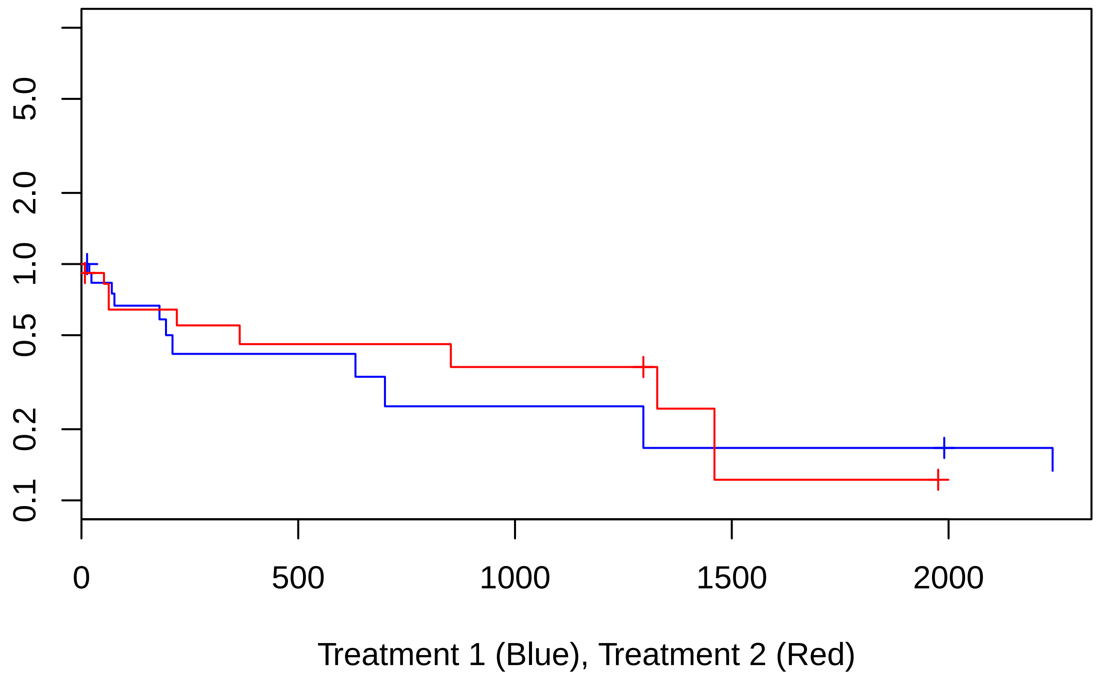
1)

```
Call: survfit(formula = Surv(G1$Time, G1$Status) ~ 1, conf.type = "log-log")

   records      n.max    n.start     events     *rmean *se(rmean)
        13         13         13         11        657        229
    median    0.95LCL    0.95UCL
       202         23       1296
    * restricted mean with upper limit =  2240
```

```
Call: survfit(formula = Surv(G1$Time, G1$Status) ~ 1, conf.type = "log-log")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   18     12       1    0.917  0.0798       0.5390        0.988
   23     11       1    0.833  0.1076       0.4817        0.956
   70     10       1    0.750  0.1250       0.4084        0.912
   76      9       1    0.667  0.1361       0.3370        0.860
  180      8       1    0.583  0.1423       0.2701        0.801
  195      7       1    0.500  0.1443       0.2085        0.736
  210      6       1    0.417  0.1423       0.1525        0.665
  632      5       1    0.333  0.1361       0.1027        0.588
  700      4       1    0.250  0.1250       0.0601        0.505
 1296      3       1    0.167  0.1076       0.0265        0.413
 2240      1       1    0.000     NaN           NA           NA
```

```
Call: survfit(formula = Surv(G2$Time, G2$Status) ~ 1, conf.type = "log-log")

   records      n.max    n.start     events     *rmean *se(rmean)
        12         12         12          9        731        216
    median    0.95LCL    0.95UCL
       365         52       1460
    * restricted mean with upper limit =  1976
```

```
Call: survfit(formula = Surv(G2$Time, G2$Status) ~ 1, conf.type = "log-log")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    8     12       1    0.917  0.0798      0.53898        0.988
   52     10       1    0.825  0.1128      0.46095        0.953
   63      9       2    0.642  0.1441      0.30225        0.848
  220      7       1    0.550  0.1499      0.23210        0.783
  365      6       1    0.458  0.1503      0.16890        0.710
  852      5       1    0.367  0.1456      0.11318        0.630
 1328      3       1    0.244  0.1392      0.04456        0.528
 1460      2       1    0.122  0.1110      0.00744        0.406
```



2) Treatment 1: Mean = 657, Median = 202 ; Treatment 2: Mean = 731, Median = 365

3) Treatment 1 seems to be more effective than treatment 2.

IV.

1) Type I censoring, the termination point of the experiment is fixed at 500 psi.

2) Left censored, we have complete data because all of the puppies already knew how to swim.

3) Uncensored, all puppies were able to learn how to swim so all data could be rec

4) Random censoring, all puppies in this group failed to learn to swim during the time of the study

V. In this case SIQR = IQR/2

$$
\begin{aligned}
r &= \tilde{\mu} - Q(.25) \\
&= Q(.75) - \tilde{\mu}
\end{aligned}
$$

$$
\begin{aligned}
SIQR &= \frac{r + \tilde{\mu} - \tilde{\mu} - r}{2} \\
&= r
\end{aligned}
$$

$$
\begin{aligned}
P\Big[Q(.25) \leq y \leq Q(.75)\Big] &= .5 \\
P\Big[\tilde{\mu} - r \leq y \leq \tilde{\mu} + r\Big] &= .5 \\
P\Big[-r \leq (y - \tilde{\mu} \leq r)\Big] &= .5 \\
P\Big[|(y - \tilde{\mu}| \leq r)\Big] &= .5 \\
r = SIQR &= |y - \tilde{\mu}|
\end{aligned}
$$