Pstat 150 Lab D

Kendall Brown

November 26, 2017

```
drugt=read.table("C:/Users/kebro/Desktop/PSTAT 105/Drug Treatment.txt",header=T)
library(survival)
```

Q1a. Looking at the data in R, I see that after day 519 most of the observations became censored with only a few outliers being observed in the 600+ day range.

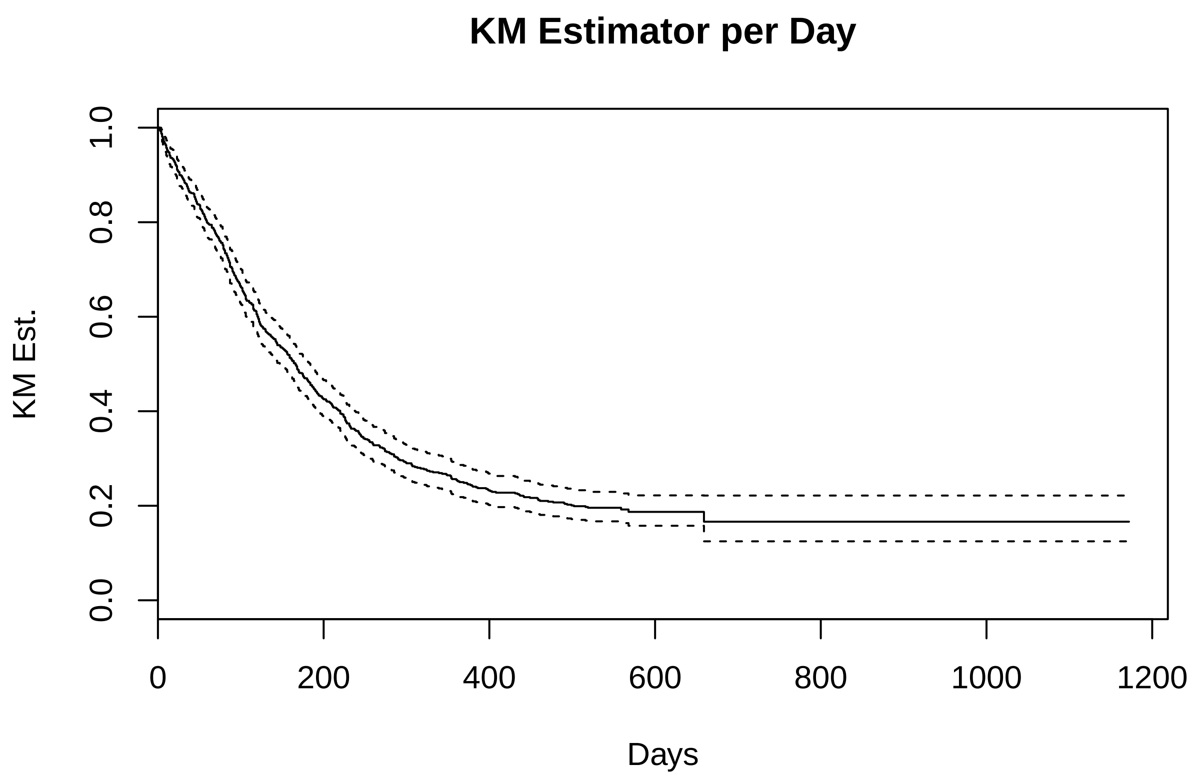
Q1b.

```
dtc=Surv(drugt$Days,drugt$Censor)
dsf=survfit(dtc~1)
dsfs=summary(dsf)
dsfs$surv
```

```
##   [1] 0.9984076 0.9936306 0.9872611 0.9808917 0.9761146 0.9697452 0.9665605
##   [8] 0.9633758 0.9570064 0.9522293 0.9490446 0.9474522 0.9410828 0.9363057
##  [15] 0.9347134 0.9331210 0.9299363 0.9267516 0.9219745 0.9187898 0.9108280
##  [22] 0.9076433 0.9060510 0.8996815 0.8980892 0.8949045 0.8917197 0.8885350
##  [29] 0.8837580 0.8821656 0.8805732 0.8757962 0.8710191 0.8662420 0.8630573
##  [36] 0.8614650 0.8598726 0.8535032 0.8487261 0.8439490 0.8391720 0.8375796
##  [43] 0.8359873 0.8280255 0.8264331 0.8232484 0.8184713 0.8168790 0.8121019
##  [50] 0.8073248 0.8041401 0.7993631 0.7977707 0.7961783 0.7945860 0.7882166
##  [57] 0.7850318 0.7818471 0.7770701 0.7738854 0.7707006 0.7691083 0.7659236
##  [64] 0.7611465 0.7595541 0.7563694 0.7515924 0.7436306 0.7404459 0.7340764
##  [71] 0.7292994 0.7245223 0.7197452 0.7149682 0.7054140 0.7022293 0.6958599
##  [78] 0.6926752 0.6878981 0.6863057 0.6815287 0.6783439 0.6751592 0.6735669
##  [85] 0.6703822 0.6656051 0.6624204 0.6608280 0.6544586 0.6512739 0.6480892
##  [92] 0.6449045 0.6369427 0.6337580 0.6305732 0.6289809 0.6257962 0.6162420
##  [99] 0.6130573 0.6114650 0.6050955 0.6003185 0.5955414 0.5891720 0.5843949
## [106] 0.5812102 0.5796178 0.5780255 0.5748408 0.5732484 0.5684713 0.5668790
## [113] 0.5652866 0.5636943 0.5621019 0.5589172 0.5573248 0.5557325 0.5541401
## [120] 0.5525478 0.5477707 0.5445860 0.5398089 0.5382166 0.5350318 0.5334395
## [127] 0.5318471 0.5302548 0.5286624 0.5270701 0.5254777 0.5207006 0.5191083
## [134] 0.5127389 0.5095541 0.5063694 0.5015924 0.4984076 0.4952229 0.4888535
## [141] 0.4872611 0.4824841 0.4808917 0.4792994 0.4745223 0.4713376 0.4697452
## [148] 0.4665605 0.4633758 0.4617834 0.4601911 0.4554140 0.4522293 0.4506369
## [155] 0.4474522 0.4458599 0.4426752 0.4410828 0.4378981 0.4363057 0.4331210
## [162] 0.4315287 0.4283439 0.4267516 0.4251592 0.4219745 0.4203822 0.4187898
## [169] 0.4171975 0.4156051 0.4124204 0.4092357 0.4076433 0.4060510 0.4044586
## [176] 0.4028662 0.4012739 0.3949045 0.3933121 0.3885350 0.3869427 0.3805732
## [183] 0.3773885 0.3742038 0.3694268 0.3662420 0.3630573 0.3614650 0.3598726
## [190] 0.3582803 0.3550955 0.3519108 0.3503185 0.3471338 0.3455414 0.3423567
## [197] 0.3407643 0.3391720 0.3359873 0.3343949 0.3312102 0.3280255 0.3264331
## [204] 0.3232484 0.3216561 0.3200637 0.3152866 0.3136943 0.3121019 0.3105096
## [211] 0.3089172 0.3041401 0.3025478 0.3009554 0.2977707 0.2961783 0.2945860
## [218] 0.2929936 0.2914013 0.2898089 0.2866242 0.2834395 0.2818471 0.2802548
## [225] 0.2786624 0.2770701 0.2754777 0.2738854 0.2722930 0.2707006 0.2691083
## [232] 0.2675159 0.2659236 0.2643312 0.2627389 0.2579618 0.2563694 0.2547771
## [239] 0.2531847 0.2515924 0.2500000 0.2484076 0.2468153 0.2452229 0.2436306
```

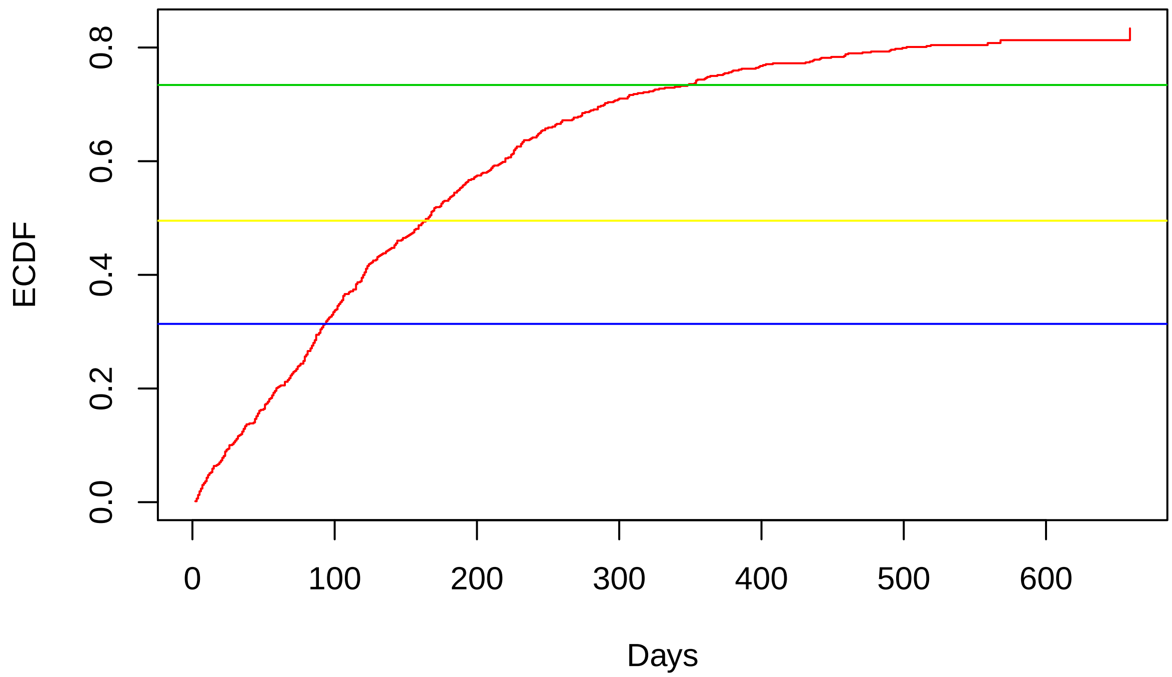
```
## [246] 0.2420382 0.2404459 0.2388535 0.2372611 0.2356688 0.2340764 0.2324841
## [253] 0.2308917 0.2292994 0.2277070 0.2261146 0.2245223 0.2229299 0.2213376
## [260] 0.2197452 0.2181529 0.2165605 0.2149682 0.2117834 0.2101911 0.2085987
## [267] 0.2070064 0.2054140 0.2038217 0.2022293 0.2006369 0.1990318 0.1973732
## [274] 0.1957146 0.1920903 0.1870353 0.1662536
```

```r
plot(dsf,main="KM Estimator per Day",xlab="Days",ylab="KM Est.")
```



Q1b.

```r
dsfecdf=1-dsfs$surv
dsqt=quantile(dsfs$time,c(.25,.5,.75))
dsqe=quantile(dsfs$surv,c(.25,.5,.75))
plot(dsfs$time,dsfecdf,type="s",xlab="Days",ylab="ECDF",col="red")
abline(h=dsqe,col=c("blue","yellow","green3"))
```




```
dsqe
```

```
##       25%       50%       75%
## 0.3136943 0.4952229 0.7340764
```

```
dsqt
```

```
## 25% 50% 75%
##  81 167 276
```

We have the .25, .5, and .75 quantiles of days to be 81, 167, and 276 respectively.

Q1d.

```
drugpre=subset(drugt,IVDrug=="Previous",data=drugt)
drugrec=subset(drugt,IVDrug=="Recent",data=drugt)
drugnev=subset(drugt,IVDrug=="Never",data=drugt)
dtpc=Surv(drugpre$Days,drugpre$Censor)
dtrc=Surv(drugrec$Days,drugrec$Censor)
dtnc=Surv(drugnev$Days,drugnev$Censor)
dspf=survfit(dtpc~1)
dsrf=survfit(dtrc~1)
dsnf=survfit(dtnc~1)
dspfs=summary(dspf)
dsrfs=summary(dsrf)
dsnfs=summary(dsnf)
dspqt=quantile(dspfs$time,c(.25,.5,.75))
dsrqt=quantile(dsrfs$time,c(.25,.5,.75))
dsnqt=quantile(dsnfs$time,c(.25,.5,.75))
```

```
##.25,.5,.75 Quantiles for the previous, recent, and never observations in terms of time
dspqt
```

```
##    25%    50%    75%
##  60.75 124.50 231.50
```

```
dsrqt
```

```
## 25% 50% 75%
##  70 136 246
```

```
dsnqt
```

```
##    25%    50%    75%
##  83.75 156.50 244.25
```

Q1e.

```
dspqe=quantile(dspfs$surv,c(.25,.5,.75))
dsrqe=quantile(dsrfs$surv,c(.25,.5,.75))
dsnqe=quantile(dsnfs$surv,c(.25,.5,.75))
dspqe[1]
```

```
##       25%
## 0.3869565
```

```
dsrqe[1]
```

```
##       25%
## 0.3091603
```

```
dsrqe[1]
```

```
##       25%
## 0.3091603
```

```
##95% Conf int for KM estimator for previous IV users after 75% mark
c(.295,.474)
```

```
## [1] 0.295 0.474
```

```
##95% Conf int for KM estimator for recent IV users after 75% mark
c(.2508,.363)
```

```
## [1] 0.2508 0.3630
```

```
##95% Conf int for KM estimator for people that have never had an IV after 75% mark
c(.255,.374)
```

```
## [1] 0.255 0.374
```

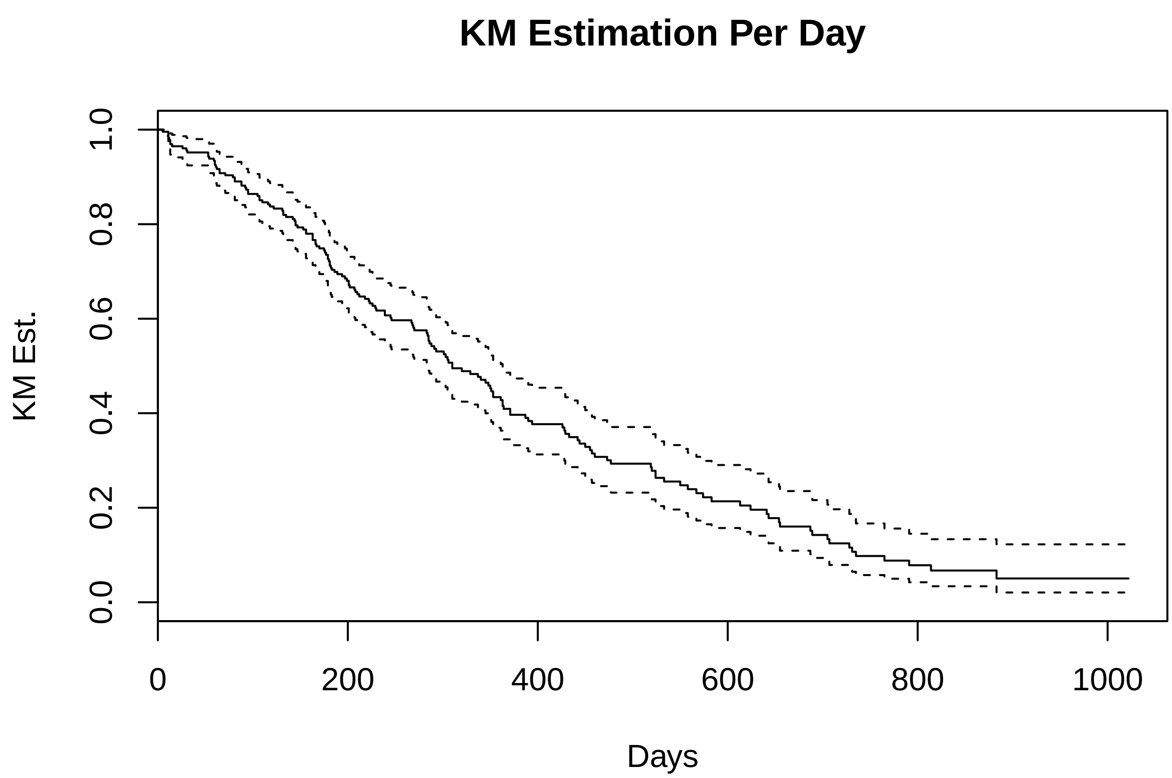
From these confidence intervals, It can be said that there does not exist a significant differance between groups. Q2.

```
lungt=read.table("C:/Users/kebro/Desktop/PSTAT 105/lung.txt",header=T)
library(survival)
```

Q2a.

```
lungsurv=Surv(lungt$time,abs(-1*lungt$status+1))
lsf=survfit(lungsurv~1)
```

```r
lsfs=summary(lsf)
plot(lsf,main="KM Estimation Per Day",xlab = "Days",ylab="KM Est.")
```



Q2b.

```r
summary(lsf)
```

```
## Call: survfit(formula = lungsurv ~ 1)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     5    228       1   0.9956 0.00438       0.9871        1.000
##    11    227       3   0.9825 0.00869       0.9656        1.000
##    12    224       1   0.9781 0.00970       0.9592        0.997
##    13    223       2   0.9693 0.01142       0.9472        0.992
##    15    221       1   0.9649 0.01219       0.9413        0.989
##    26    220       1   0.9605 0.01290       0.9356        0.986
##    30    219       1   0.9561 0.01356       0.9299        0.983
##    31    218       1   0.9518 0.01419       0.9243        0.980
##    53    217       2   0.9430 0.01536       0.9134        0.974
##    54    215       1   0.9386 0.01590       0.9079        0.970
##    59    214       1   0.9342 0.01642       0.9026        0.967
##    60    213       2   0.9254 0.01740       0.8920        0.960
##    61    211       1   0.9211 0.01786       0.8867        0.957
##    62    210       1   0.9167 0.01830       0.8815        0.953
##    65    209       2   0.9079 0.01915       0.8711        0.946
##    71    207       1   0.9035 0.01955       0.8660        0.943
##    79    206       1   0.8991 0.01995       0.8609        0.939
```

```
##    81    205       2   0.8904 0.02069       0.8507        0.932
##    88    203       2   0.8816 0.02140       0.8406        0.925
##    92    201       1   0.8772 0.02174       0.8356        0.921
##    93    199       1   0.8728 0.02207       0.8306        0.917
##    95    198       2   0.8640 0.02271       0.8206        0.910
##   105    196       1   0.8596 0.02302       0.8156        0.906
##   107    194       2   0.8507 0.02362       0.8056        0.898
##   110    192       1   0.8463 0.02391       0.8007        0.894
##   116    191       1   0.8418 0.02419       0.7957        0.891
##   118    190       1   0.8374 0.02446       0.7908        0.887
##   122    189       1   0.8330 0.02473       0.7859        0.883
##   131    188       1   0.8285 0.02500       0.7810        0.879
##   132    187       2   0.8197 0.02550       0.7712        0.871
##   135    185       1   0.8153 0.02575       0.7663        0.867
##   142    184       1   0.8108 0.02598       0.7615        0.863
##   144    183       1   0.8064 0.02622       0.7566        0.859
##   145    182       2   0.7975 0.02667       0.7469        0.852
##   147    180       1   0.7931 0.02688       0.7421        0.848
##   153    179       1   0.7887 0.02710       0.7373        0.844
##   156    178       2   0.7798 0.02751       0.7277        0.836
##   163    176       3   0.7665 0.02809       0.7134        0.824
##   166    173       2   0.7577 0.02845       0.7039        0.816
##   167    171       1   0.7532 0.02863       0.6991        0.811
##   170    170       1   0.7488 0.02880       0.6944        0.807
##   175    167       1   0.7443 0.02898       0.6896        0.803
##   176    165       1   0.7398 0.02915       0.6848        0.799
##   177    164       1   0.7353 0.02932       0.6800        0.795
##   179    162       2   0.7262 0.02965       0.6704        0.787
##   180    160       1   0.7217 0.02981       0.6655        0.783
##   181    159       2   0.7126 0.03012       0.6559        0.774
##   182    157       1   0.7081 0.03027       0.6511        0.770
##   183    156       1   0.7035 0.03041       0.6464        0.766
##   186    154       1   0.6989 0.03056       0.6416        0.761
##   189    152       1   0.6943 0.03070       0.6367        0.757
##   194    149       1   0.6897 0.03085       0.6318        0.753
##   197    147       1   0.6850 0.03099       0.6269        0.749
##   199    145       1   0.6803 0.03113       0.6219        0.744
##   201    144       2   0.6708 0.03141       0.6120        0.735
##   202    142       1   0.6661 0.03154       0.6071        0.731
##   207    139       1   0.6613 0.03168       0.6020        0.726
##   208    138       1   0.6565 0.03181       0.5970        0.722
##   210    137       1   0.6517 0.03194       0.5920        0.717
##   212    135       1   0.6469 0.03206       0.5870        0.713
##   218    134       1   0.6421 0.03218       0.5820        0.708
##   222    132       1   0.6372 0.03231       0.5769        0.704
##   223    130       1   0.6323 0.03243       0.5718        0.699
##   226    126       1   0.6273 0.03256       0.5666        0.694
##   229    125       1   0.6223 0.03268       0.5614        0.690
##   230    124       1   0.6172 0.03280       0.5562        0.685
##   239    121       2   0.6070 0.03304       0.5456        0.675
##   245    117       1   0.6019 0.03316       0.5402        0.670
##   246    116       1   0.5967 0.03328       0.5349        0.666
##   267    112       1   0.5913 0.03341       0.5294        0.661
##   268    111       1   0.5860 0.03353       0.5239        0.656
```

```
##   269     110        1   0.5807 0.03364       0.5184        0.651
##   270     108        1   0.5753 0.03376       0.5128        0.645
##   283     104        1   0.5698 0.03388       0.5071        0.640
##   284     103        1   0.5642 0.03400       0.5014        0.635
##   285     101        2   0.5531 0.03424       0.4899        0.624
##   286      99        1   0.5475 0.03434       0.4841        0.619
##   288      98        1   0.5419 0.03444       0.4784        0.614
##   291      97        1   0.5363 0.03454       0.4727        0.608
##   293      94        1   0.5306 0.03464       0.4669        0.603
##   301      91        1   0.5248 0.03475       0.4609        0.597
##   303      89        1   0.5189 0.03485       0.4549        0.592
##   305      87        1   0.5129 0.03496       0.4488        0.586
##   306      86        1   0.5070 0.03506       0.4427        0.581
##   310      85        2   0.4950 0.03523       0.4306        0.569
##   320      82        1   0.4890 0.03532       0.4244        0.563
##   329      81        1   0.4830 0.03539       0.4183        0.558
##   337      79        1   0.4768 0.03547       0.4121        0.552
##   340      78        1   0.4707 0.03554       0.4060        0.546
##   345      77        1   0.4646 0.03560       0.3998        0.540
##   348      76        1   0.4585 0.03565       0.3937        0.534
##   350      75        1   0.4524 0.03569       0.3876        0.528
##   351      74        1   0.4463 0.03573       0.3815        0.522
##   353      73        2   0.4340 0.03578       0.3693        0.510
##   361      70        1   0.4278 0.03581       0.3631        0.504
##   363      69        2   0.4154 0.03583       0.3508        0.492
##   364      67        1   0.4092 0.03582       0.3447        0.486
##   371      65        2   0.3966 0.03581       0.3323        0.473
##   387      60        1   0.3900 0.03582       0.3258        0.467
##   390      59        1   0.3834 0.03582       0.3193        0.460
##   394      58        1   0.3768 0.03580       0.3128        0.454
##   426      55        1   0.3700 0.03580       0.3060        0.447
##   428      54        1   0.3631 0.03579       0.2993        0.440
##   429      53        1   0.3563 0.03576       0.2926        0.434
##   433      52        1   0.3494 0.03573       0.2860        0.427
##   442      51        1   0.3426 0.03568       0.2793        0.420
##   444      50        1   0.3357 0.03561       0.2727        0.413
##   450      48        1   0.3287 0.03555       0.2659        0.406
##   455      47        1   0.3217 0.03548       0.2592        0.399
##   457      46        1   0.3147 0.03539       0.2525        0.392
##   460      44        1   0.3076 0.03530       0.2456        0.385
##   473      43        1   0.3004 0.03520       0.2388        0.378
##   477      42        1   0.2933 0.03508       0.2320        0.371
##   519      39        1   0.2857 0.03498       0.2248        0.363
##   520      38        1   0.2782 0.03485       0.2177        0.356
##   524      37        2   0.2632 0.03455       0.2035        0.340
##   533      34        1   0.2554 0.03439       0.1962        0.333
##   550      32        1   0.2475 0.03423       0.1887        0.325
##   558      30        1   0.2392 0.03407       0.1810        0.316
##   567      28        1   0.2307 0.03391       0.1729        0.308
##   574      27        1   0.2221 0.03371       0.1650        0.299
##   583      26        1   0.2136 0.03348       0.1571        0.290
##   613      24        1   0.2047 0.03325       0.1489        0.281
##   624      23        1   0.1958 0.03297       0.1407        0.272
##   641      22        1   0.1869 0.03265       0.1327        0.263
```

```
##   643     21       1   0.1780 0.03229       0.1247        0.254
##   654     20       1   0.1691 0.03188       0.1169        0.245
##   655     19       1   0.1602 0.03142       0.1091        0.235
##   687     18       1   0.1513 0.03090       0.1014        0.226
##   689     17       1   0.1424 0.03034       0.0938        0.216
##   705     16       1   0.1335 0.02972       0.0863        0.207
##   707     15       1   0.1246 0.02904       0.0789        0.197
##   728     14       1   0.1157 0.02830       0.0716        0.187
##   731     13       1   0.1068 0.02749       0.0645        0.177
##   735     12       1   0.0979 0.02660       0.0575        0.167
##   765     10       1   0.0881 0.02568       0.0498        0.156
##   791      9       1   0.0783 0.02462       0.0423        0.145
##   814      7       1   0.0671 0.02351       0.0338        0.133
##   883      4       1   0.0503 0.02285       0.0207        0.123
```

```
##95% conf int of KM estimator for 150 days
c(.7421,.848)
```

```
## [1] 0.7421 0.8480
```

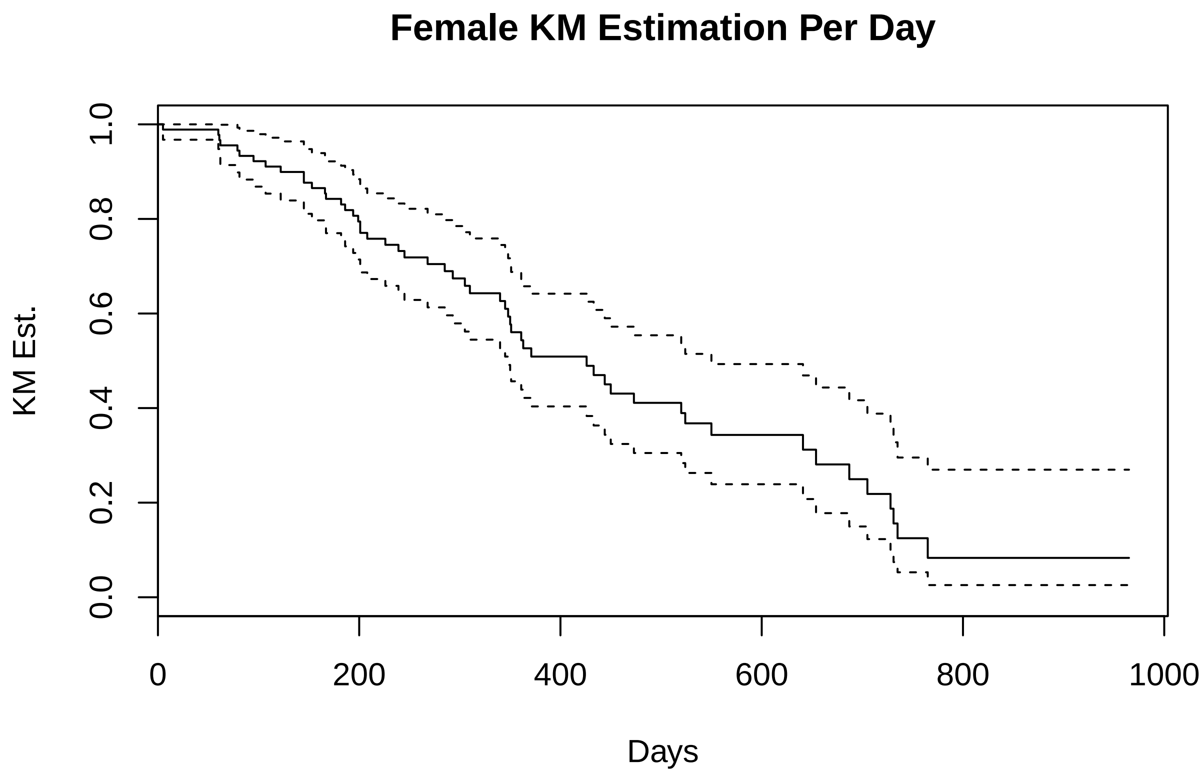
Q2c.

```
##95% conf int of median survival time
lsf
```

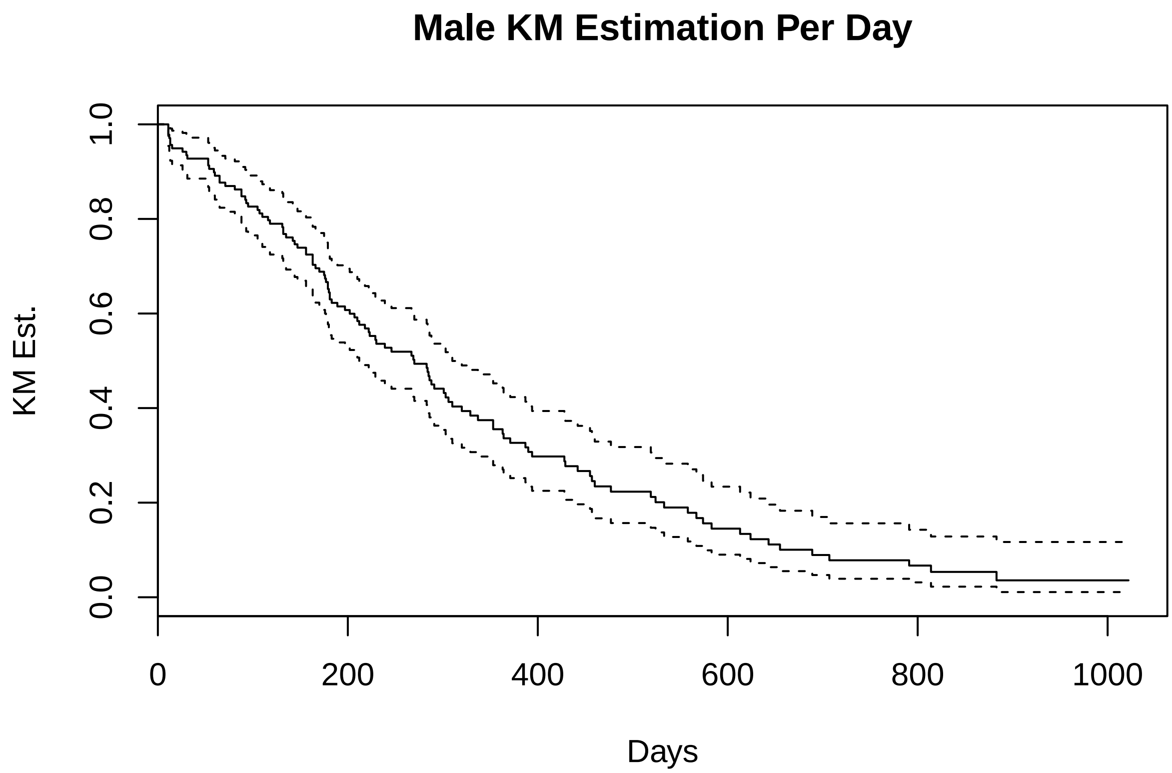
```
## Call: survfit(formula = lungsurv ~ 1)
##
##       n  events  median 0.95LCL 0.95UCL
##     228     165     310     285     363
```

we have a 95% conf int of {285,363} Q2d.

```
lungmale=subset(lungt,lungt$sex==1)
lungfemale=subset(lungt,lungt$sex==2)
lungmalesf=Surv(lungmale$time,abs(-1*lungmale$status+1))
lungfemalesf=Surv(lungfemale$time,abs(-1*lungfemale$status+1))
lungmfs=survfit(lungmalesf~1)
lungffs=survfit(lungfemalesf~1)
plot(lungffs,main="Female KM Estimation Per Day",xlab = "Days",ylab="KM Est.")
```




```
plot(lungmfs,main="Male KM Estimation Per Day",xlab = "Days",ylab="KM Est.")
```



From these plot it appears that women have higher survival rates seemingly along the entire

Q2e. MvF median surv times

```
lungmfs
```

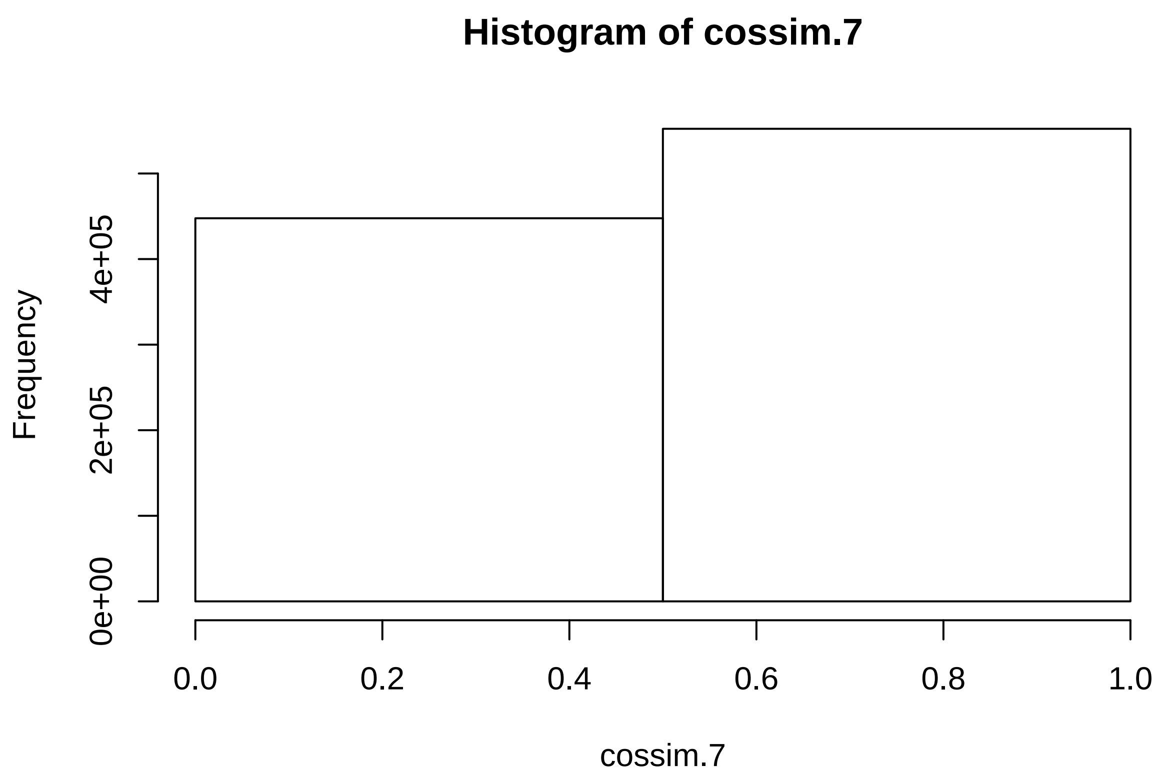
```
## Call: survfit(formula = lungmalesf ~ 1)
##
##       n  events  median 0.95LCL 0.95UCL
##     138     112     270     212     310
```

```
lungffs
```

```
## Call: survfit(formula = lungfemalesf ~ 1)
##
##       n  events  median 0.95LCL 0.95UCL
##      90      53     426     348     550
```

From these confidence intervals and medians, I believe it is rather clear that women tend to survive longer than men. However, i do not believe that this tells the whole story as other factors not measured in this data set could be rather impactful to survival time. Q3.1,000,000 sims

```
simexp=rexp(10^6)
cossim=cos(simexp)
cossim.7=(cossim>.7)/1
cossim.7h=hist(cossim.7,breaks=c(0,.5,1))
```




```
cossim.7p=cossim.7h$counts[2]/(10^6)
cossim.7p ##Probability Cos(x)>.7
```

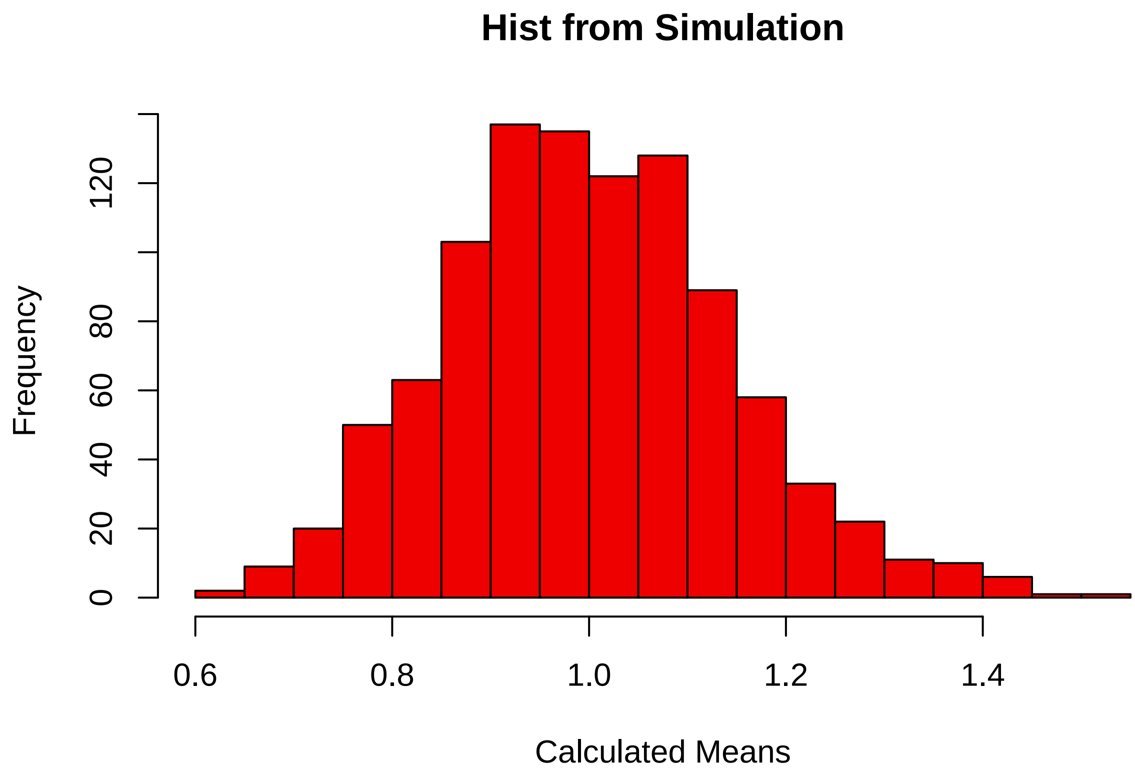
```
## [1] 0.552281
```

```
##95% conf int for P(cos(x)>.7)
c(cossim.7p-1.96*sqrt(cossim.7p*(1-cossim.7p)/(10^6)),
  cossim.7p+1.96*sqrt(cossim.7p*(1-cossim.7p)/(10^6)))
```

```
## [1] 0.5513064 0.5532556
```

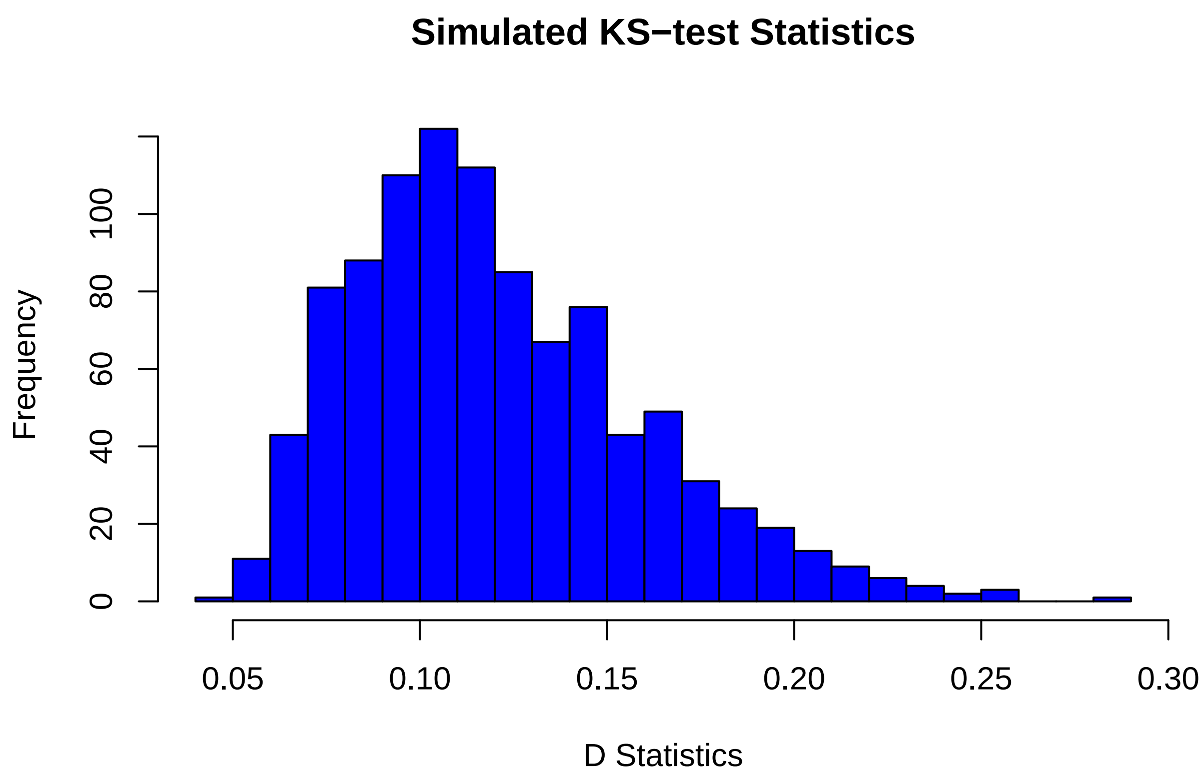
Q4a.1000 runs of 50 sims

```
n=50
m=1000
ex.test=rep(999,m)
rnd.x=matrix(rep(999,50000),1000,50)
for(j in 1:m){
  rnd.x[j,]=c(rexp(n))
  ex.test[j]=mean(rnd.x[j,])
}
hist(ex.test,breaks=30,col="red2",main="Hist from Simulation",xlab="Calculated Means")
```



Q4b. KS-test of samples

```
kssim=rep(999,1000)
for(j in 1:m){
  ksdj=ks.test(rnd.x[j,],pexp)
  kssim[j]=ksdj$statistic
}
histks=hist(kssim,breaks=30,main="Simulated KS-test Statistics",xlab="D Statistics",col="blue")
```



Q4c.

```
quantile(kssim,.95)
```

```
##       95%
## 0.1936762
```

Q4d.

```
#computer ran a single batch in .16 seconds
#7200 secs in 2hrs means 45000 batches
#to avoid time issues when knitting this block will be ## out

#kssim2h=rep(999,m)
#rnd2h.x=matrix(rep(999,50000),1000,50)
#batch=matrix(rep(999,4.5e+07),45000,1000)
#for(j in 1:45000){
#for(k in 1:m){
  #rnd2h.x[k,]=c(rexp(n))
  #ex.test[k]=mean(rnd2h.x[k,])
  #ksdj2h=ks.test(rnd2h.x[k,],pexp)
  #kssim2h[k]=ksdj2h$statistic
  #batch[j,]=kssim2h
#}}
```

```
#sim2h=quantile(batch,.95)
#sim2h
```

To avoid problems during knitting, I am stating the calculated simulation .95 quantile to be .1884515.

Q4e.Crit value from Stephens table 1.4

```
(.1884515-.2/50)*(sqrt(50)+.26+.5/sqrt(50))
```

```
## [1] 1.365269
```

From this calculation we get a T value that corresponds to the <1% level.

Q5a.

```
library(MASS)
ofsd=sd(geyser$waiting)
ofsd ##sample standard deviation of geyser data set
```

```
## [1] 13.89032
```

Q5b.

```
geysermat=rep(999,1000)
for(j in 1:1000){
sampof=sample(geyser$waiting,299,replace = T)
geysermat[j]=sd(sampof)
}
(sd(geysermat))^2
```

```
## [1] 0.1501207
```

Q5c.

```
geyseriqr=rep(999,1000)
for(j in 1:1000){
sampofiqr=sample(geyser$waiting,299,replace = T)
geyseriqr[j]=IQR(sampofiqr)
}
(sd(geyseriqr))^2
```

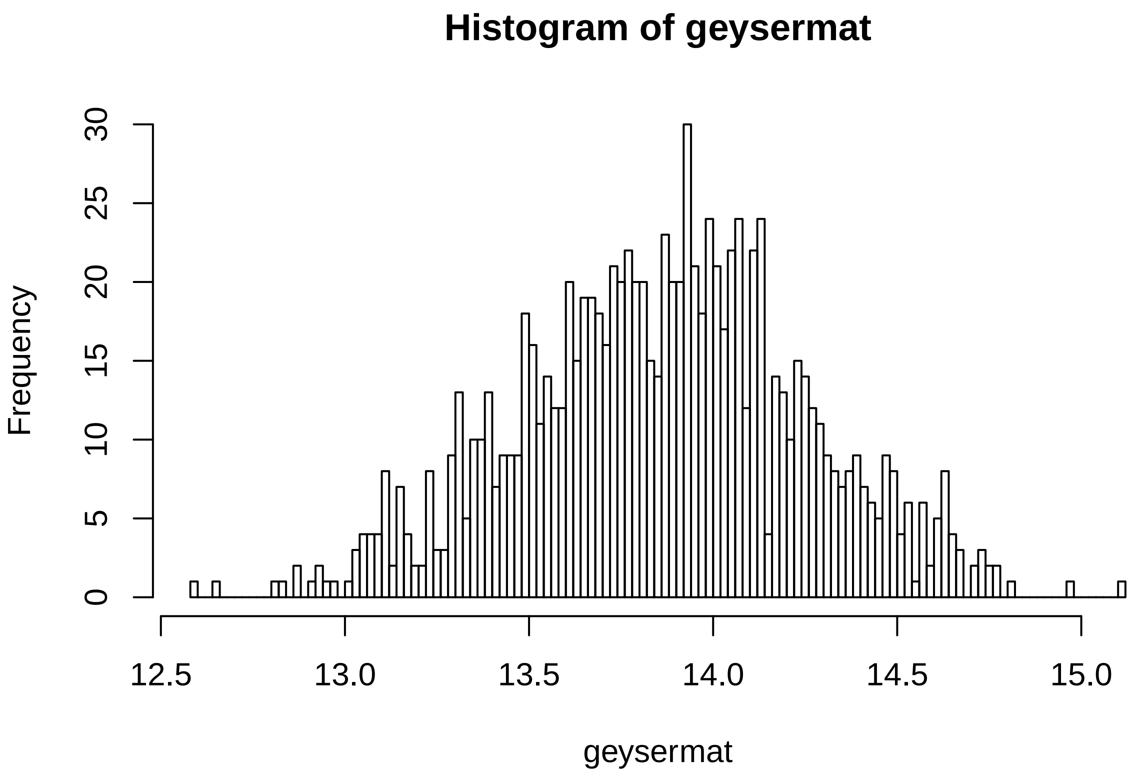
```
## [1] 2.863103
```

Q5d.

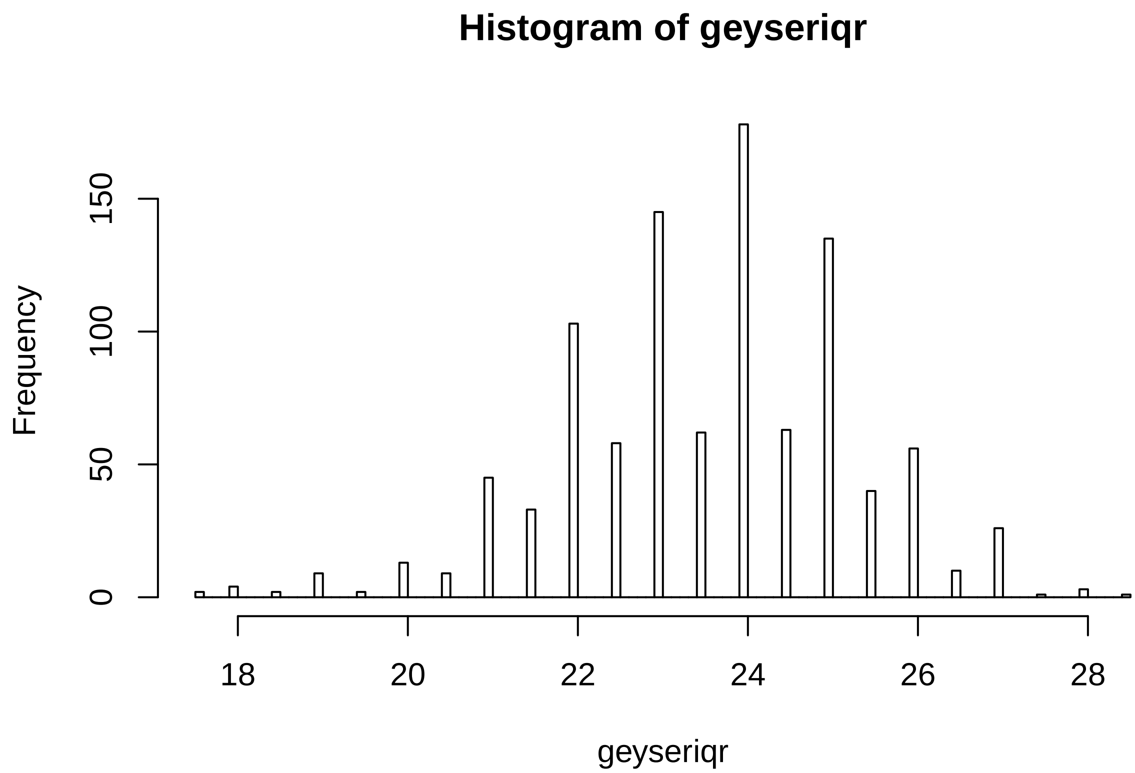
```
geysermad=rep(999,1000)
for(j in 1:1000){
sampofmad=sample(geyser$waiting,299,replace = T)
geysermad[j]=mad(sampofmad)
}
(sd(geysermad))^2
```

```
## [1] 2.965954
```

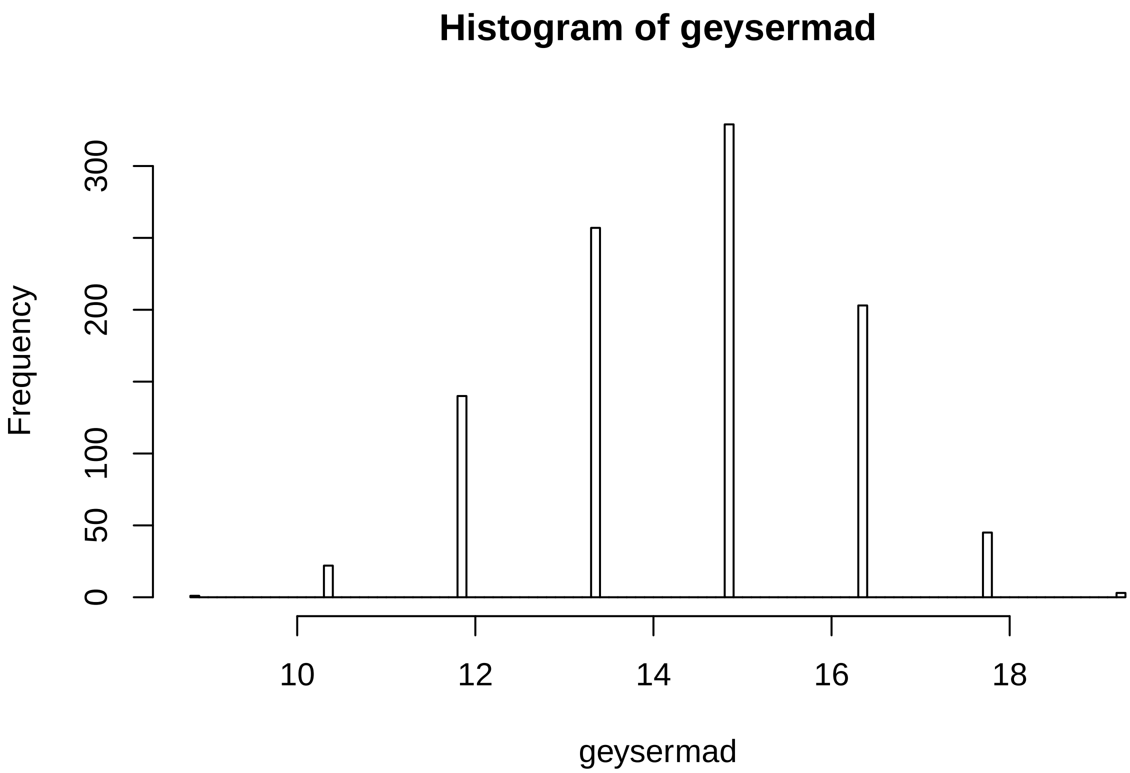
```
hist(geysermat,breaks=100)
```




```
hist(geyseriqr,breaks=100)
```




```
hist(geysermad,breaks=100)
```



Based on the histograms of the standard deviation estimates, the SD function has the most normaly distributed data. As such, personally, i prefer the sd function over the other two estimation functions.