P8131_hw9

Shan Jiang

Problem 1

Determine the survival and density functions for a continuous survival time variable with hazard function: $\lambda(x) = h(x) = \frac{2x}{(1+x^2)}$

$$S(t) = exp\{-\int_0^t \lambda(x)dx\} = exp\{-\int_0^t \frac{2x}{(1+x^2)}dx\} = exp\{-ln(1+t^2)\} = -(1+t^2)$$

$f(t) = S(t) \times \lambda(t) = exp\{-ln(1+t^2)\} \times \frac{2t}{(1+t^2)} = 2t$

Problem 2

(1) K_M estimate of survival function

```
ti <- rbind(1, 2, 4, 5, 6, 7, 8, 9, 10)
ni <- rbind(10, 9, 7, 6, 5, 4, 3, 2, 1)
di <- rbind(1, 2, 0, 0, 1, 0, 0, 0, 0)
ci <- rbind(0, 0, 1, 1, 0,  1, 1, 1, 1)
Hi <- rbind(di/ni) ## K_M estimate of survival function

S1 = 1 * (1- Hi[1,])
S2 = S1 * 1 * (1- Hi[2,])

St = function(t) {
        1 * (1- Hi[t,])
}

hat_st = rbind(1 * St(1),
               St(1) * St(2),
               St(1) * St(2) * St(3),
               St(1) * St(2) * St(3) * St(4),
               St(1) * St(2) * St(3) * St(4)* St(5),
               St(1) * St(2) * St(3) * St(4) * St(5) * St(6),
               St(1) * St(2) * St(3) * St(4)* St(5) * St(6) * St(7),
               St(1) * St(2) * St(3) * St(4)* St(5) * St(6) * St(7) * St(8),
               St(1) * St(2) * St(3) * St(4)*  St(5) * St(6) * St(7) * St(8) * St(9))


lifetable <- data.frame(ti = ti,
                ni = ni,
                di = di,
                ci = ci,
                Hi = Hi, St = hat_st)

knitr::kable(lifetable, digits = 2)
```

ti	ni	di	ci	Hi	St
1	10	1	0	0.10	0.90
2	9	2	0	0.22	0.70
4	7	0	1	0.00	0.70
5	6	0	1	0.00	0.70
6	5	1	0	0.20	0.56
7	4	0	1	0.00	0.56
8	3	0	1	0.00	0.56
9	2	0	1	0.00	0.56
10	1	0	1	0.00	0.56

The K_M estimate of survival function is defined as
$\hat{S}(t) = \prod_{i=1}^{k}(1 - \hat{H}_i)$, KM estimator is $\hat{H}_t = -log(\hat{S}_t)$

$$\hat{S}(1) = \prod_{i=1}^{1}(1 - \hat{H}_i) = 1 \times (1 - 0.1) = 0.9$$

$$\hat{S}(2) = \prod_{i=1}^{2}(1 - \hat{H}_i) = 1 \times (1 - 0.1) \times (1 - \frac{2}{9}) = 0.7$$

$$\hat{S}(4) = \prod_{i=1}^{3}(1 - \hat{H}_i) = 1 \times (1 - 0.1) \times (1 - \frac{2}{9}) \times (1 - \frac{0}{7}) = 0.7$$

$$\hat{S}(5) = \prod_{i=1}^{4}(1 - \hat{H}_i) = 1 \times (1 - 0.1) \times (1 - \frac{2}{9}) \times (1 - \frac{0}{7}) \times (1 - \frac{0}{6}) = 0.7$$

$$\hat{S}(6) = \prod_{i=1}^{5}(1 - \hat{H}_i) = 1 \times (1 - 0.1) \times (1 - \frac{2}{9}) \times (1 - \frac{0}{7}) \times (1 - \frac{0}{6}) \times (1 - \frac{1}{5}) = 0.56$$

$\hat{S}(7) = \prod_{i=1}^{7}(1 - \hat{H}_i) = 1 \times (1 - 0.1) \times (1 - \frac{2}{9}) \times (1 - \frac{0}{7}) \times (1 - \frac{0}{6}) \times (1 - \frac{1}{5}) \times (1 - \frac{0}{4}) = 0.56$

$$\hat{S}(8) = \prod_{i=1}^{8}(1 - \hat{H}_i) = 1 \times (1 - 0.1) \times (1 - \frac{2}{9}) \times (1 - \frac{0}{7}) \times (1 - \frac{0}{6}) \times (1 - \frac{1}{5}) \times (1 - \frac{0}{4}) \times (1 - \frac{0}{3}) = 0.56$$

$$\hat{S}(9) = \prod_{i=1}^{9}(1 - \hat{H}_i) = 1 \times (1 - 0.1) \times (1 - \frac{2}{9}) \times (1 - \frac{0}{7}) \times (1 - \frac{0}{6}) \times (1 - \frac{1}{5}) \times (1 - \frac{0}{4}) \times (1 - \frac{0}{3}) \times (1 - \frac{0}{2}) = 0.56$$

$$\hat{S}(10) = \prod_{i=1}^{10}(1 - \hat{H}_i) = 1 \times (1 - 0.1) \times (1 - \frac{2}{9}) \times (1 - \frac{0}{7}) \times (1 - \frac{0}{6}) \times (1 - \frac{1}{5}) \times (1 - \frac{0}{4}) \times (1 - \frac{0}{3}) \times (1 - \frac{0}{2}) \times (1 - \frac{0}{1}) = 0.56$$

(2) N__A estimate of cumulative hazard function;

- Nelson-Aalen estimator:

$$t_1 = \sum_{t_i \le t} \frac{d_i}{n_i} = \frac{d_1}{n_1} = \frac{1}{10} = 0.1$$

$$t_2 = \sum_{t_i \le t} \frac{d_i}{n_i} = \frac{d_1}{n_1} + \frac{d_2}{n_2} = \frac{2}{9} + \frac{1}{10} = \frac{11}{90}$$

$$t_4 = \sum_{t_i \le t} \frac{d_i}{n_i} = \frac{d_1}{n_1} + \frac{d_2}{n_2} + \frac{d_3}{n_3} = \frac{2}{9} + \frac{1}{10} + \frac{0}{7} = \frac{11}{90}$$

$$t_5 = \sum_{t_i \le t} \frac{d_i}{n_i} = \frac{d_1}{n_1} + \cdots + \frac{d_5}{n_5} = \frac{2}{9} + \frac{1}{10} + \frac{0}{7} + \frac{0}{6} = \frac{11}{90}$$

$$t_6 = \sum_{t_i \le t} \frac{d_i}{n_i} = \frac{d_1}{n_1} + \cdots + \frac{d_6}{n_6} = \frac{2}{9} + \frac{1}{10} + \frac{0}{7} + \frac{0}{6} + \frac{1}{5} == \frac{29}{90}$$

$$t_7 = \sum_{t_i \le t} \frac{d_i}{n_i} = \frac{d_1}{n_1} + \cdots + \frac{d_7}{n_7} = \frac{2}{9} + \frac{1}{10} + \frac{0}{7} + \frac{0}{6} + \frac{1}{5} + \frac{0}{4} = \frac{29}{90}$$

$$t_8 = \sum_{t_i \le t} \frac{d_i}{n_i} = \frac{d_1}{n_1} + \cdots + \frac{d_8}{n_8} = \frac{2}{9} + \frac{1}{10} + \frac{0}{7} + \frac{0}{6} + \frac{1}{5} + \frac{0}{4} + \frac{0}{3} = \frac{29}{90}$$

$$t_9 = \sum_{t_i \le t} \frac{d_i}{n_i} = \frac{d_1}{n_1} + \cdots + \frac{d_9}{n_9} = \frac{2}{9} + \frac{1}{10} + \frac{0}{7} + \frac{0}{6} + \frac{1}{5} + \frac{0}{4} + \frac{0}{3} + \frac{0}{2} = \frac{29}{90}$$

$$t_{10} = \sum_{t_i \le t} \frac{d_i}{n_i} = \frac{d_1}{n_1} + \cdots + \frac{d_{10}}{n_{10}} = \frac{2}{9} + \frac{1}{10} + \frac{0}{7} + \frac{0}{6} + \frac{1}{5} + \frac{0}{4} + \frac{0}{3} + \frac{0}{2} + \frac{0}{1} = \frac{29}{90}$$

(3) Fleming-Harrington estimate of survival function

Fleming-Harrington estimate is $exp(-\tilde{H}(t))$ for S(t):

- $exp(-\tilde{H}(1)) = exp(-(t_1)) = exp(-0.1) \approx 0.9048$
- $exp(-\tilde{H}(2)) = exp(-(t_2)) = exp(-0.122) \approx 0.8850$
- $exp(-\tilde{H}(4)) = exp(-0.122) \approx 0.8850$
- $exp(-\tilde{H}(5)) = exp(-0.122) \approx 0.8850$
- $exp(-\tilde{H}(6))exp(-0.322) \approx 0.7246$
- $exp(-\tilde{H}(7)) \approx 0.7246$
- $exp(-\tilde{H}(8)) \approx 0.7246$
- $exp(-\tilde{H}(9)) \approx 0.7246$
- $exp(-\tilde{H}(10)) \approx 0.7246$

Problem 3

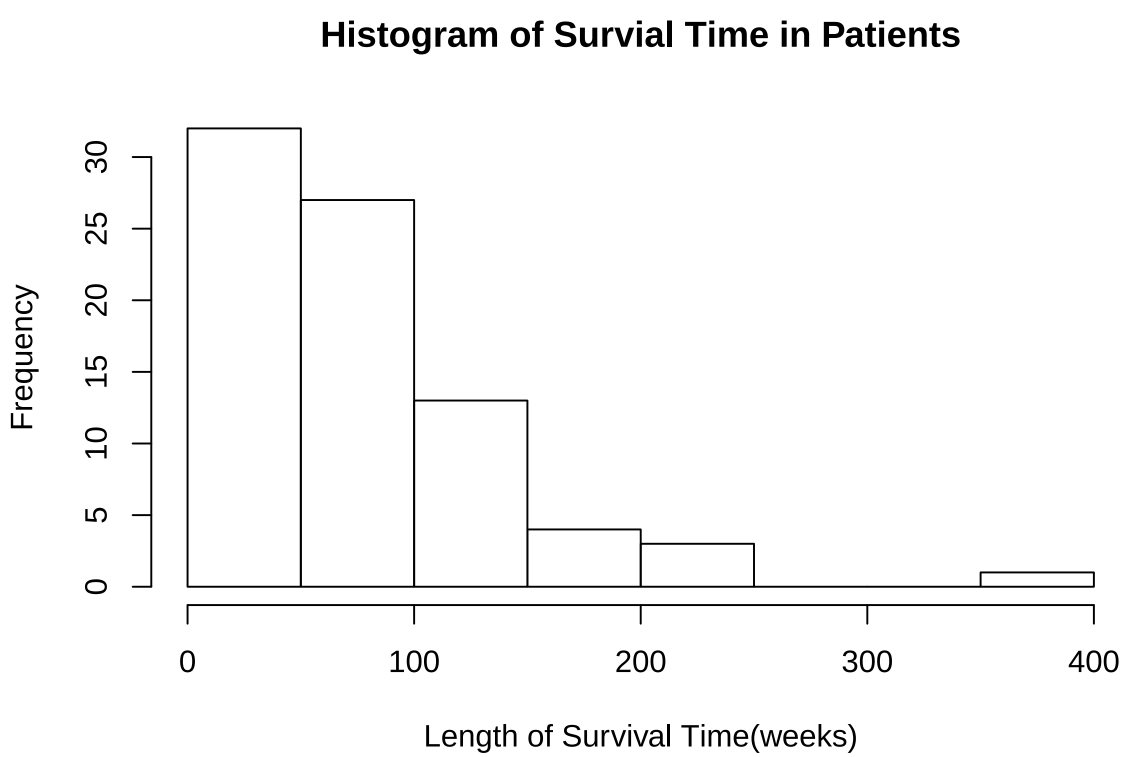
This data frame contains the following columns:

type Tumor DNA profile (1=Aneuploid Tumor, 2=Diploid Tumor)

time Time to death or on-study time, weeks

delta Death indicator (0=alive, 1=dead)

```r
## load the data
library(KMsurv)
library(survival)
data(tongue)

## histogram for overview of data
hist(tongue$time, xlab="Length of Survival Time(weeks)", main="Histogram of Survial Time in Patients")
```

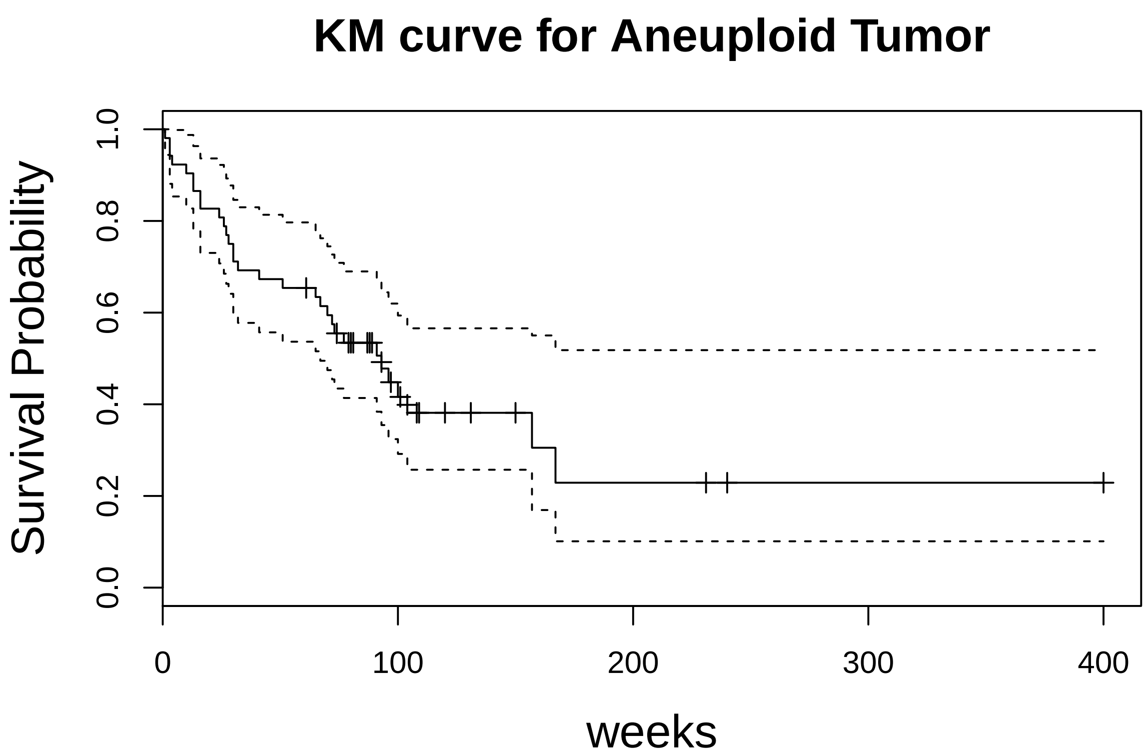


Plot the KM-curve

Aneuploid Tumor

```r
# KM survival function -Aneuploid Tumor
KM1 = survfit(Surv(time, delta)~1,
              data = subset(tongue, type =="1"),
```

```r
             conf.type = 'log')

## K_M Curve for Aneuploid Tumor
plot(KM1, conf.int = T, mark.time = TRUE,
     xlab="weeks", ylab = "Survival Probability",
     main="KM curve for Aneuploid Tumor", cex.lab=1.5, cex.main=1.5)
```




```r
summary(KM1)
```

```
## Call: survfit(formula = Surv(time, delta) ~ 1, data = subset(tongue,
##     type == "1"), conf.type = "log")
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1     52       1    0.981  0.0190        0.944        1.000
##     3     51       2    0.942  0.0323        0.881        1.000
##     4     49       1    0.923  0.0370        0.853        0.998
##    10     48       1    0.904  0.0409        0.827        0.988
##    13     47       2    0.865  0.0473        0.777        0.963
##    16     45       2    0.827  0.0525        0.730        0.936
##    24     43       1    0.808  0.0547        0.707        0.922
##    26     42       1    0.788  0.0566        0.685        0.908
##    27     41       1    0.769  0.0584        0.663        0.893
##    28     40       1    0.750  0.0600        0.641        0.877
##    30     39       2    0.712  0.0628        0.598        0.846
```

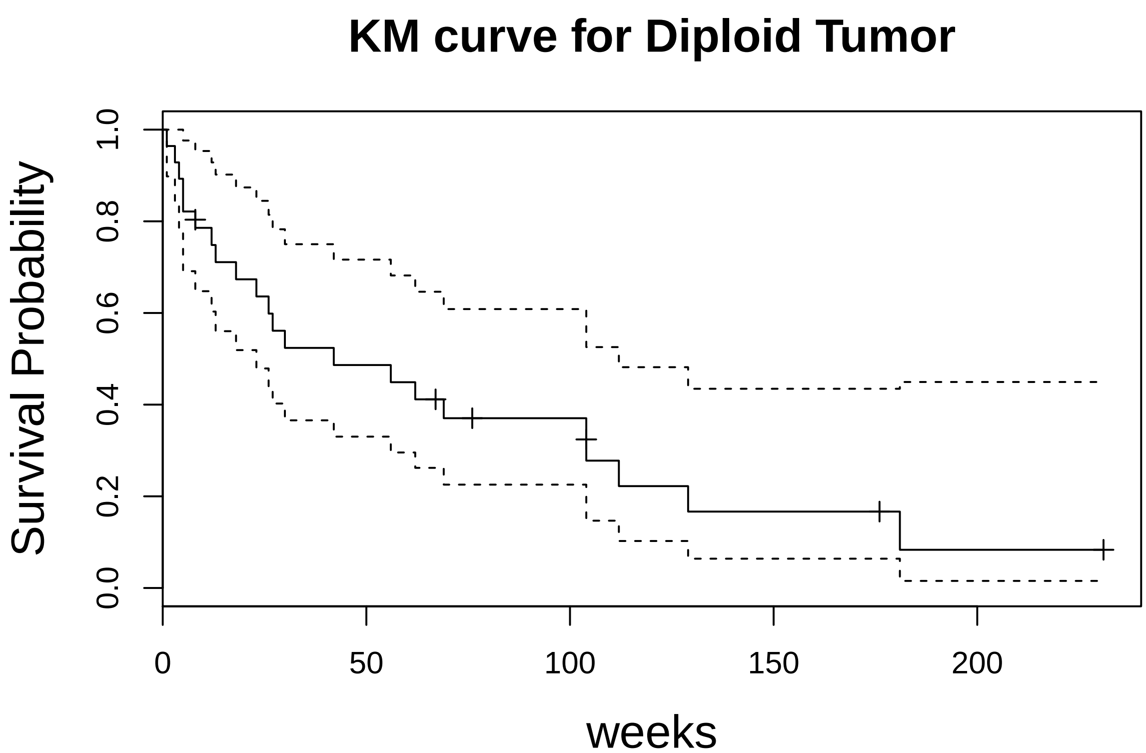
```
##    32     37        1    0.692  0.0640        0.578         0.830
##    41     36        1    0.673  0.0651        0.557         0.813
##    51     35        1    0.654  0.0660        0.537         0.797
##    65     33        1    0.634  0.0669        0.516         0.780
##    67     32        1    0.614  0.0677        0.495         0.762
##    70     31        1    0.594  0.0683        0.475         0.745
##    72     30        1    0.575  0.0689        0.454         0.727
##    73     29        1    0.555  0.0693        0.434         0.709
##    77     27        1    0.534  0.0697        0.414         0.690
##    91     19        1    0.506  0.0715        0.384         0.667
##    93     18        1    0.478  0.0728        0.355         0.644
##    96     16        1    0.448  0.0741        0.324         0.620
##   100     14        1    0.416  0.0754        0.292         0.594
##   104     12        1    0.381  0.0767        0.257         0.566
##   157      5        1    0.305  0.0918        0.169         0.550
##   167      4        1    0.229  0.0954        0.101         0.518
```

Diploid Tumor

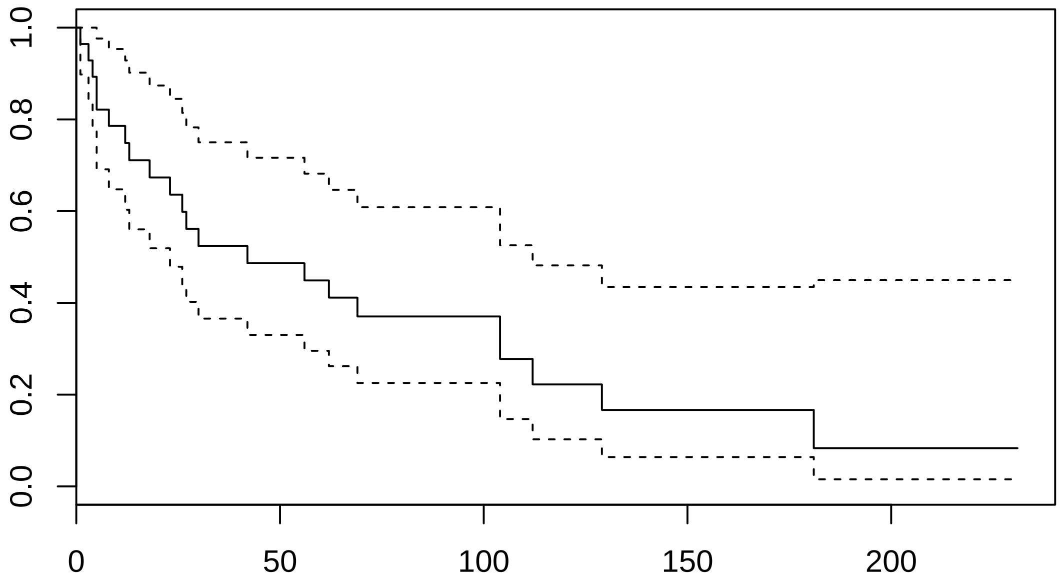
```r
# KM survival function -Diploid Tumor
KM2 = survfit(Surv(time, delta)~1,
              data = subset(tongue, type =="2"),
              conf.type = 'log')


summary(KM2)
```

```
## Call: survfit(formula = Surv(time, delta) ~ 1, data = subset(tongue,
##     type == "2"), conf.type = "log")
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1     28       1   0.9643  0.0351       0.8979        1.000
##     3     27       1   0.9286  0.0487       0.8379        1.000
##     4     26       1   0.8929  0.0585       0.7853        1.000
##     5     25       2   0.8214  0.0724       0.6911        0.976
##     8     23       1   0.7857  0.0775       0.6475        0.953
##    12     21       1   0.7483  0.0824       0.6031        0.929
##    13     20       1   0.7109  0.0863       0.5603        0.902
##    18     19       1   0.6735  0.0895       0.5190        0.874
##    23     18       1   0.6361  0.0921       0.4790        0.845
##    26     17       1   0.5986  0.0939       0.4402        0.814
##    27     16       1   0.5612  0.0952       0.4025        0.783
##    30     15       1   0.5238  0.0959       0.3658        0.750
##    42     14       1   0.4864  0.0961       0.3302        0.716
##    56     13       1   0.4490  0.0957       0.2956        0.682
##    62     12       1   0.4116  0.0948       0.2621        0.646
##    69     10       1   0.3704  0.0938       0.2255        0.608
##   104      8       2   0.2778  0.0904       0.1468        0.526
##   112      5       1   0.2222  0.0877       0.1025        0.482
##   129      4       1   0.1667  0.0815       0.0639        0.435
##   181      2       1   0.0833  0.0717       0.0155        0.449
```

```r
plot(KM2, conf.int = T, mark.time = TRUE,
     xlab="weeks", ylab = "Survival Probability",
     main="KM curve for Diploid Tumor", cex.lab=1.5, cex.main=1.5)
```




```r
plot(KM2)
```



Est. One-year survival rate and 95% CI

One-year is 52 weeks

```r
# obtain survival rate at given time, with CI for Type I: 0.654
summary(KM1,time = c(52))
```

```
## Call: survfit(formula = Surv(time, delta) ~ 1, data = subset(tongue,
##     type == "1"), conf.type = "log")
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    52     34      18    0.654   0.066        0.537        0.797
```

The one-year(52 weeks) survival rate is 0.654, with 95% CI as [0.537, 0.797] for Aneuploid Tumor.

```r
# obtain survival rate at given time, with CI for Type II: 0.486
summary(KM2,time = c(52))
```

```
## Call: survfit(formula = Surv(time, delta) ~ 1, data = subset(tongue,
##     type == "2"), conf.type = "log")
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    52     13      14    0.486  0.0961         0.33        0.716
```

The one-year(52 weeks) survival rate is 0.486, with 95% CI as of [0.33, 0.716] for Diploid Tumor.