30/4/2024

Μοντέλα Αξιοπιστίας και Επιβίωσης

ΣΕΙΡΑ 2

Κοντοθανάση Σωτηρία
ΑΜ: 09120080

(A) Γράψτε τη λογαριθμοποιημένη συνάρτηση πιθανοφάνειας για το μοντέλο της Λογαριθμο–κανονικής κατανομής όταν έχουμε δεξιά αποκομμένες παρατηρήσεις.

Γνωρίζουμε ότι η σππ της Λογαριθμο–κανονικής δίνεται από τον τύπο:

$$f(t) = \frac{1}{\sigma t\sqrt{2\pi}} exp\left(\frac{-(lnt-\mu)^2}{2\sigma^2}\right), \qquad t > 0$$

Γνωρίζουμε ότι η συνάρτηση επιβίωσης της Λογαριθμο–κανονικής δίνεται από τον τύπο:

$$S(t) \;=\; 1 - \Phi\left(\frac{lnt-\mu}{\sigma}\right)$$

Επιπλέον γνωρίζουμε ότι για αποκομμένες παρατηρήσεις ισχύει :

$$L = \prod_{i\in U} f(t_i) \prod_{i\in C} S(t_i) = \prod_{i} \{f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}\}$$

Επομένως έχουμε ότι :

$$L = \prod_{i\in U}\left\{\frac{1}{\sigma t\sqrt{2\pi}} exp\left(\frac{-(lnt-\mu)^2}{2\sigma^2}\right)\right\} \prod_{i\in C}\left\{1 - \Phi\left(\frac{lnt-\mu}{\sigma}\right)\right\}$$

και

$$\ell = \ln(L) = \sum\nolimits_{i\in U} \ln\left\{\frac{1}{\sigma t\sqrt{2\pi}} e^{\left(\frac{-(lnt-\mu)^2}{2\sigma^2}\right)}\right\} \sum\nolimits_{i\in C} ln\left\{1 - \Phi\left(\frac{lnt-\mu}{\sigma}\right)\right\}$$

$$= \sum\nolimits_{i\in U}\left\{\ln\left(\frac{1}{\sigma t\sqrt{2\pi}}\right) + \frac{-(lnt-\mu)^2}{2\sigma^2}\right\} \sum\nolimits_{i\in C} ln\left\{1 - \Phi\left(\frac{lnt-\mu}{\sigma}\right)\right\}$$

$$= \sum\nolimits_{i\in U}\left\{-\ln\left(\sigma t\sqrt{2\pi}\right) + \frac{-(lnt-\mu)^2}{2\sigma^2}\right\} + \sum\nolimits_{i\in C} ln\left\{1 - \Phi\left(\frac{lnt-\mu}{\sigma}\right)\right\}$$

$$= -\boldsymbol{k}\ln\left(\sigma t\sqrt{2\pi}\right) + \boldsymbol{k}\frac{-(lnt-\mu)^2}{2\sigma^2} + \sum\nolimits_{i\in C} ln\left\{1 - \Phi\left(\frac{lnt-\mu}{\sigma}\right)\right\}$$

όπου $\boldsymbol{k}$ το πλήθος των τιμών χωρίς αποκοπή .

(B) Ο παρακάτω πίνακας δίνει τη διάρκεια ζωής (σε ημέρες) 20 εξαρτημάτων, μετά από συνεχή χρήση.

725	838	853	965	1139	1142	1304	1317	1427
1658	1764	1776	1990	2010	2224	2244*	2279*	2286*

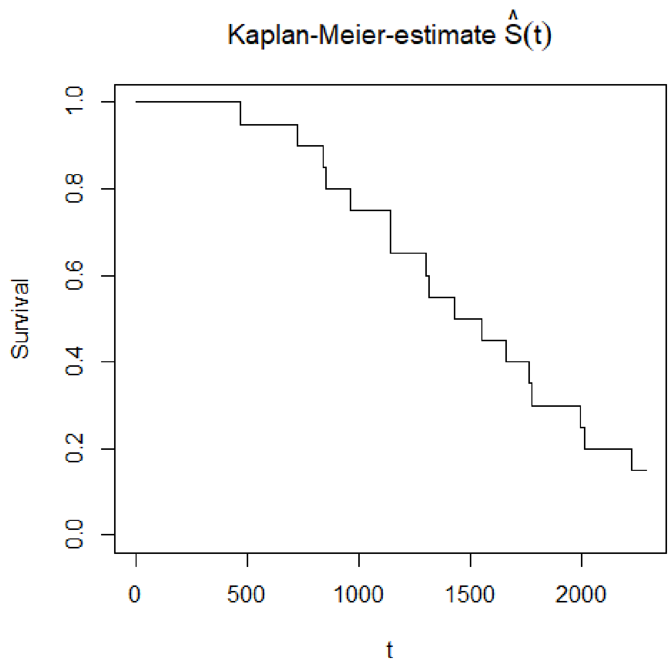
(∗=δεξιά αποκομμένες παρατηρήσεις)

i. Να βρεθούν οι εκτιμήσεις Kaplan–Meier και να γίνει η γραφική παράσταση αυτών.

Εισαγωγή των δεδομένων στην R :

```
> cbdat <-
read.table("C:\\Users\\Admin\\Desktop\\8oΕΞΑΜΗΜΟ\\data-
seira2ExB.txt", header = TRUE)
> attach(cbdat)
The following object is masked from cbdat (pos = 3):

    t

The following object is masked from cbdat (pos = 4):

    t

The following object is masked from cbdat (pos = 5):

    t

The following objects are masked from cbdat (pos = 6):

    censor, t

> cbdat
      t censor
1   468      1
2   725      1
3   838      1
4   853      1
5   965      1
6  1139      1
7  1142      1
8  1304      1
9  1317      1
10 1427      1
11 1554      1
12 1658      1
13 1764      1
14 1776      1
15 1990      1
16 2010      1
17 2224      1
18 2244      0
19 2279      0
20 2286      0
```

```
> outp<-survfit(Surv(t, censor)~1)
> summary(outp)
Call: survfit(formula = Surv(t, censor) ~ 1)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
  468     20       1     0.95  0.0487       0.8591        1.000
  725     19       1     0.90  0.0671       0.7777        1.000
  838     18       1     0.85  0.0798       0.7071        1.000
  853     17       1     0.80  0.0894       0.6426        0.996
  965     16       1     0.75  0.0968       0.5823        0.966
 1139     15       1     0.70  0.1025       0.5254        0.933
 1142     14       1     0.65  0.1067       0.4712        0.897
 1304     13       1     0.60  0.1095       0.4195        0.858
 1317     12       1     0.55  0.1112       0.3700        0.818
 1427     11       1     0.50  0.1118       0.3226        0.775
 1554     10       1     0.45  0.1112       0.2772        0.731
 1658      9       1     0.40  0.1095       0.2339        0.684
 1764      8       1     0.35  0.1067       0.1926        0.636
 1776      7       1     0.30  0.1025       0.1536        0.586
 1990      6       1     0.25  0.0968       0.1170        0.534
 2010      5       1     0.20  0.0894       0.0832        0.481
 2224      4       1     0.15  0.0798       0.0528        0.426

> plot(outp, main=expression(paste("Kaplan-Meier-estimate ",
hat(S)(t))),
+ conf.int=FALSE,xlab="t", ylab="Survival")
```



Παρατηρήσεις:

Παρατηρώ ότι όσο ο χρόνος περνάει , όλο και λιγότερες μονάδες βρίσκονται σε κίνδυνο. Επιπλέον παρατηρώ ότι με την πάροδο του χρόνου όλο και περισσότερες μονάδες αποτυγχάνουν – χαλάνε (failed).

Όσον αφορά την γραφική παράσταση της Kaplan – Meier, η οποία εκτιμά με μη παραμετρικό τρόπο την συνάρτηση επιβίωσης , παρατηρώ ότι τα σκαλάκια δεν πέφτουν απότομα προς τα κάτω αλλά είναι πιο εξαπλώμενα, που σημαίνει ότι τα γεγονότα δεν συμβαίνουν γρηγορα στην πάροδο του χρονου. Έχω δηλαδη ένα φαινομενο που εξιλεσσεται αργα με την πάροδο του χρόνου.

ii. **Κάνοντας χρήση των εκτιμήσεων Kaplan–Meier Sˆ(ti) να γίνουν γραφικοί έλεγχοι για τις κατανομές: Εκθετική, Weibull, Λογαριθμο–κανονική και Λογαριθμο–λογιστική.**

Εξετάζω αν τα δεδομένα μου ακολουθούν την Εκθετική Κατανομή .

Αρχικά βρίσκω τη λίστα των εκτιμημένων πιθανοτήτων επιβίωσης με χρήση της Kaplan - Meier για κάθε χρονικό σημείο όπως έχω υπολογίσει προηγούμενος.

```
> outp$surv
 [1] 0.95 0.90 0.85 0.80 0.75 0.70 0.65 0.60 0.55 0.50 0.45 0.40
0.35 0.30 0.25
[16] 0.20 0.15 0.15 0.15 0.15
```

Έπειτα βρίσκω τις εκτιμημένες πιθανότητες επιβίωσης με χρήση της Kaplan - Meier μόνο για τα χρονικά σημεία όπου συνέβη ένα συμβάν (δηλαδή **n.event == 1**).

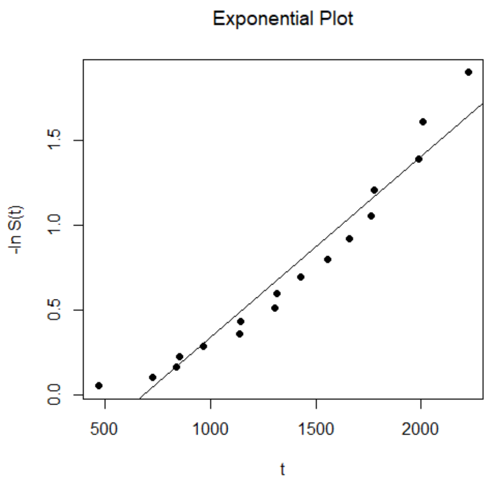
```
> SKM<-outp$surv [outp$n.event ==1]
> SKM
 [1] 0.95 0.90 0.85 0.80 0.75 0.70 0.65 0.60 0.55 0.50 0.45 0.40
0.35 0.30 0.25
[16] 0.20 0.15
```

Ύστερα βρίσκω τα χρονικά σημεία κατά τα οποία συνέβη ένα συμβάν (δηλαδή **n.event == 1**).

```
> Utime<-outp$time[outp$n.event ==1]
> Utime
 [1]  468  725  838  853  965 1139 1142 1304 1317 1427 1554 1658
1764 1776 1990
[16] 2010 2224
```

Τελος , κάνοντας χρήση των εκτιμησεων της Kaplan Meier και με κατασκευη του παρακάτω γραφηματος θα δω εάν τα δεδομενα του πειραματος μας, για τα οποία συνέβη το γεγονος ακολουθούν το μοντέλο της Εκθετικης Κατονομης, .

```
> ##Examine whether data follows Exponential distribution
>
>     plot(Utime,-log(SKM),     main=expression(paste("Exponential
Plot")), xlab="t",
+ ylab="-ln S(t)", pch=19)
> abline(lm(-log(SKM)~Utime))
```



Συμπέρασμα :

Σύμφωνα με το διαγγραμμα στα δεξία, λόγω μικρής τάυτισης των αποτελεσματων(εκτιμησεων ΚΜ) με την ευθεία έχω αρκετές ενδείξης ώστε να απορριψω την υποθέση , ότι τα δεδομενα ακολουθούν το μοντέλο της Εκθετική Κατανομή.

Εξετάζω αν τα δεδομένα μου ακολουθούν την Κατανομή Weibull .

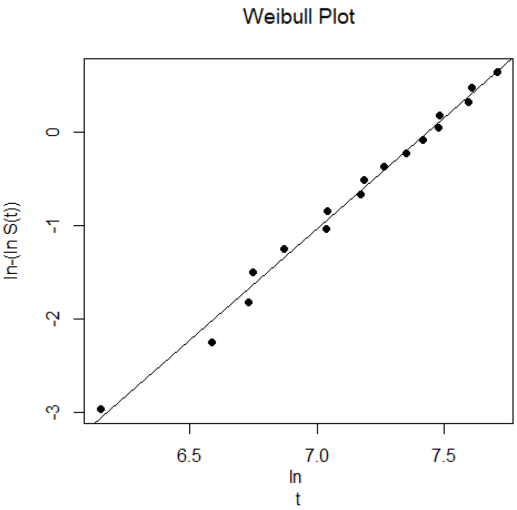
Σε αυτήν την περίπτωση, κάνοντας χρήση των εκτιμησεων της Kaplan Meier και με κατασκευη του παρακάτω γραφηματος θα δω εάν τα δεδομενα του πειραματος μας, για τα οποία συνέβη το γεγονος ακολουθούν το μοντέλο της Κατονομης Weibull .

```
> ##Examine whether data follows Weibull distribution
>
>  plot(log(Utime),log(-log(SKM)),  main=expression(paste("Weibull
Plot")), xlab="ln
+ t", ylab="ln-(ln S(t))", pch=19)
> abline(lm(log(-log(SKM))~log(Utime)))
```

Συμπέρασμα :

Σύμφωνα με το διαγγραμμα στα δεξία, λόγω μεγάλης τάυτισης των αποτελεσματων(εκτιμησεων KM) με την ευθεία δεν έχω αρκετές ενδείξης ώστε να απορριψω την υποθέση , ότι τα δεδομενα ακολουθουν το μοντέλο της κατανομη Weibull.

Σε αντιθεση με τον προηγουμενο έλεγχο, όπου είχα αρκετες ενδείξεις ώστε να απόρριψω την Εκθετικη Κατανομη.



Εξετάζω αν τα δεδομένα μου ακολουθούν την Κατανομή LogNormal.

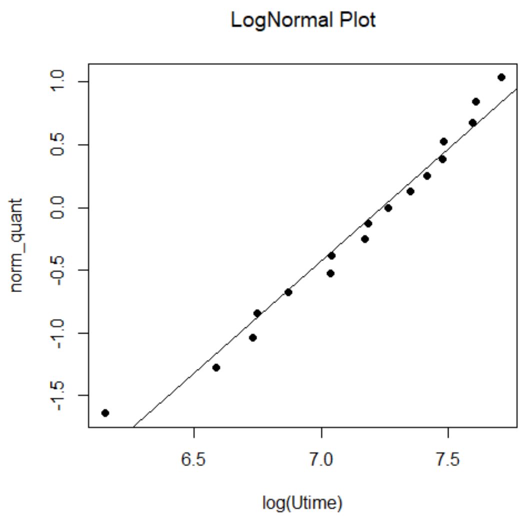
Σε αυτήν την περίπτωση, κάνοντας χρήση των εκτιμησεων της Kaplan Meier και με κατασκευη του παρακάτω γραφηματος θα δω εάν τα δεδομενα του πειραματος μας, για τα οποία συνέβη το γεγονος ακολουθούν το μοντέλο της Κατονομης LogNormal .

```
> ##Examine whether data follows LogNormal distribution
>
> norm_quant<-qnorm(1-SKM,0,1)
> norm_quant
 [1]  -1.6448536  -1.2815516  -1.0364334  -0.8416212  -0.6744898  -
0.5244005
 [7]  -0.3853205  -0.2533471  -0.1256613   0.0000000   0.1256613
0.2533471
[13]  0.3853205  0.5244005  0.6744898  0.8416212  1.0364334
> plot(log(Utime), norm_quant, pch=19)
> abline(lm(norm_quant ~ log(Utime)))
```

Συμπέρασμα :

Σύμφωνα με το διαγγραμμα στα δεξία, λόγω μεγάλης τάυτισης των αποτελεσματων(εκτιμησεων KM) με την ευθεία δεν έχω αρκετές ενδείξης ώστε να απορριψω την υποθέση , ότι τα δεδομενα ακολουθουν το μοντέλο της κατανομη Log Normal.

Άρα βρισκόμαστε σε ένα σημείο όπου τα δεδομενα του πειραματος θα μπορούσαν να ακολουθουν και το μοντελο της Weibull και το μοντέλο της LogNormal . Ωστοσο φαίνεται τα δεδομένα μου να έχουν κατά ελάχιστα μεγαλύτερη ταύτιση με την ευθεία στο μοντέλο της Weibull και μικρότερη ταύτιση με την ευθεία στο μοντέλο της LogNormal.



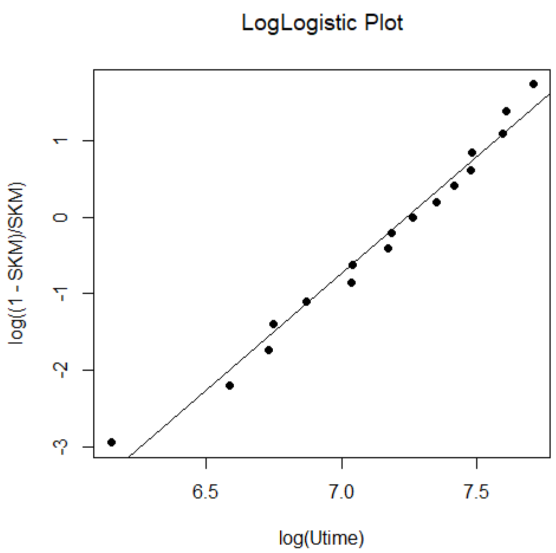
Εξετάζω αν τα δεδομένα μου ακολουθούν την <u>Κατανομή LogLogistic</u>

```
> ##Examine whether data follows LogLogistic distribution
>
> plot(log(Utime),log((1-SKM)/SKM), pch=19,,
main=expression(paste("LogLogistic Plot")))
> abline(lm(log((1-SKM)/SKM) ~ log(Utime)))
>
```

<u>Συμπέρασμα :</u>

Σύμφωνα με το διαγγραμμα στα δεξία, λόγω μεγάλης τάυτισης των αποτελεσματων(εκτιμησεων KM) με την ευθεία δεν έχω αρκετές ενδείξης ώστε να απορριψω την υποθέση , ότι τα δεδομενα ακολουθουν το μοντέλο της κατανομη Log Logistic.

Επομένως βρισκόμαστε σε ένα σημείο όπου τα δεδομενα του πειραματος θα μπορουσαν να ακολουθουν και το μοντελο της Weibull και το μοντέλο της Log Normal και το μοντέλο της Log Logistic. Ωστοσο φαίνεται τα δεδομένα μου να έχουν κατά ελάχιστα μεγαλύτερη ταύτιση με την ευθεία στο μοντέλο της Weibull και μικρότερη ταύτιση με την ευθεία στο μοντέλο της Log Normal και της Log Logistic.



iii. **Στη συνέχεια προσαρμόστε στα δεδομένα, τα παραπάνω τέσσερα μοντέλα κατανομών με τη μέθοδο μέγιστης πιθανοφάνειας (ML) και κατασκευάστε γραφικές παραστάσεις των $\widehat{S}(t_i\,;\,\widehat{\theta})\ -\ ML$ με τα σημεία $\widehat{S}(t_i) - Kaplan - Meier$ για τις κατανομές αυτές.**

Για την προσαρμογή των παρακάτω μοντέλων θα χρησιμοποιήσουμε από τα δεδομένα τον χρόνο επιβίωσης, δηλαδή τις ημέρες όπου κατάφεραν και δούλεψαν/λειτουργήσαν τα εξαρτήματα και το ποιες παρατηρήσεις ήταν δεξιά αποκομμένες και ποιες όχι, δηλαδή ποια και πόσα εξαρτήματα χάλασαν , σε ποια χρονική στιγμή και εξαρτήματα συνέχισαν να δουλεύουν.

Εκθετική Κατανομή

Πάμε αρχικά να προσαρμόσουμε το μοντέλο της Εκθετικής Κατανομής

```
> mod1<-survreg(Surv(t, censor)~1, data=cbdat, dist="exponential")
> mod1
Call:
survreg(formula = Surv(t, censor) ~ 1, data = cbdat, dist =
"exponential")

Coefficients:
(Intercept)
   7.474505

Scale fixed at 1

Loglik(model)= -144.1   Loglik(intercept only)= -144.1
n= 20
```

Παραπάνω δίνεται μία εκτίμηση της παραμέτρου α μέσω της μεθόδου μεγίστης πιθανοφάνειας.

Συμπέρασμα :

1. Συντελεστής Intercept: Ο εκτιμηθείς συντελεστής Intercept είναι περίπου 7.4745. Ο συντελεστής Intercept για την εκθετική κατανομή αναπαριστά το λογάριθμο της αναμενόμενης επιβίωσης για τον παράμετρο κλίμακας (scale parameter) της κατανομής.

2. Κλίμακα: Η κλίμακα είναι προσδιορισμένη στην τιμή 1. Αυτό σημαίνει ότι η αναλλοίωτη κλίμακα της εκθετικής κατανομής έχει οριστεί ως 1.

3. Λογαριθμική Πιθανοφάνεια (Log Likelihood): Η λογαριθμική πιθανοφάνεια για το μοντέλο είναι περίπου -144.1, ενώ η λογαριθμική πιθανοφάνεια για ένα μοντέλο που περιλαμβάνει μόνο τη σταθερά Intercept είναι επίσης περίπου -144.1. Αυτό υποδηλώνει ότι η προσθήκη οποιασδήποτε άλλης μεταβλητής στο μοντέλο δεν βελτιώνει την προσαρμογή του.

Με βάση αυτά τα αποτελέσματα, μπορούμε να συμπεράνουμε ότι η εκθετική κατανομή φαίνεται να είναι μια καλή προσαρμογή στα δεδομένα σας, αλλά δεν υπάρχει αρκετή ενδεικτική διακύμανση στις παρατηρήσεις για να βελτιώσει το μοντέλο.

Κατανομή Weibull

Τώρα προσαρμόζουμε το μοντέλο της Κατανομής Weibull.

```
>modweib<-fit_data(data = cbdat, dist = "weibull",    time = "t", censor = "censor")
> summary(modweib)
Fitting of the distribution ' weibull ' By maximum likelihood on censored data
Parameters
         estimate  Std. Error
shape    2.576253   0.5226988
scale 1752.966512 166.4985523
Loglikelihood:  -136.6462   AIC:  277.2924   BIC:  279.2839
Correlation matrix:
         shape    scale
shape 1.000000 0.129957
scale 0.129957 1.000000
```

Συμπέρασμα :

1. **Παράμετροι Κατανομής Weibull:**

Η εκτιμηθείσα τιμή του παραμέτρου shape είναι περίπου 2.576253, που υποδηλώνει το σχήμα της κατανομής Weibull. Μια τιμή του shape μεγαλύτερη από 1 υποδεικνύει μια καμπύλη πιθανοφάνειας που αυξάνεται, ενώ μια τιμή μικρότερη από 1 υποδεικνύει μια καμπύλη πιθανοφάνειας που μειώνεται.

Η εκτιμηθείσα τιμή της παραμέτρου scale είναι περίπου 1752.966512, που αντιστοιχεί στην κλίμακα της κατανομής Weibull. Η κλίμακα καθορίζει το χρονικό διάστημα στο οποίο παρατηρούνται οι συμβάντα.

2. **Παράμετροι Εκτίμησης:** Η τυπική απόκλιση (Std. Error) για την εκτίμηση του shape είναι περίπου 0.5226988 και για την εκτίμηση της κλίμακας είναι περίπου 166.4985523.

3. **Στατιστικά Κριτήρια:**

Η λογαριθμική πιθανοφάνεια (Loglikelihood) για το μοντέλο είναι περίπου -136.6462.

Το AIC είναι περίπου 277.2924.

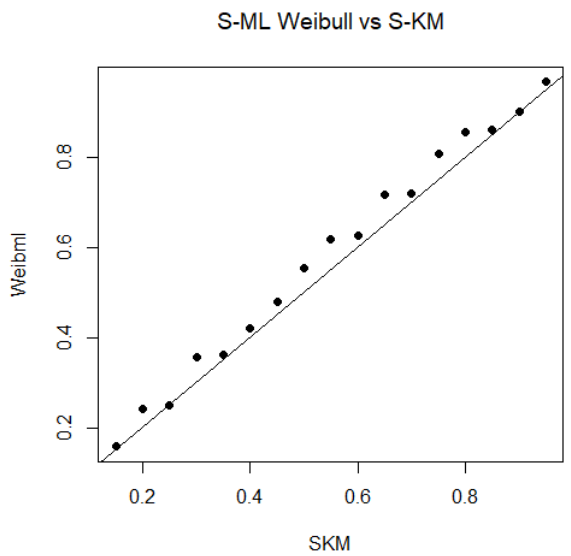
Το BIC είναι περίπου 279.2839.

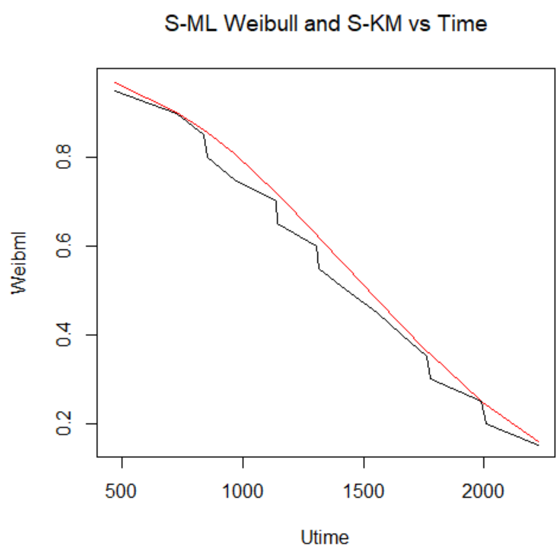
4. **Συσχέτιση Παραμέτρων:** Η συσχέτιση μεταξύ των παραμέτρων shape και scale είναι περίπου 0.129957. Αυτό υποδεικνύει το επίπεδο της συσχέτισης μεταξύ των δύο παραμέτρων.

Συνολικά, τα αποτελέσματα αυτά δείχνουν ότι η κατανομή Weibull προσαρμόζεται καλά στα δεδομένα σας, με τις εκτιμηθείσες παραμέτρους να παρέχουν μια καλή περιγραφή της καμπύλης επιβίωσης.

Στη συνέχεια , θα προχωρήσω στην κατασκευή γραφικής παράστασης S-ML vs S-KM.

```
> #1.
> outp<-survfit(Surv(t,censor)~1)
> SKM<-outp$surv [outp$n.event ==1]
> SKM
 [1] 0.95 0.90 0.85 0.80 0.75 0.70 0.65 0.60 0.55 0.50 0.45 0.40
0.35 0.30 0.25
[16] 0.20 0.15
>
> Utime<-outp$time[outp$n.event ==1]
> Utime
 [1]  468  725  838  853  965 1139 1142 1304 1317 1427 1554 1658
1764 1776 1990
[16] 2010 2224
>
>
> #2.
> Weibml<- pweibull(Utime, shape= modweib$estimate[1], scale =
modweib$estimate[2],
+ lower.tail = FALSE, log.p = FALSE)
> Weibml
 [1] 0.9672481 0.9022690 0.8612589 0.8552638 0.8066701 0.7194243
0.7178152
 [8] 0.6271194 0.6195902 0.5551095 0.4803845 0.4204924 0.3619331
0.3555097
[15] 0.2499656 0.2410829 0.1578320
>
> #3.
> plot(Weibml~SKM, pch=19, main=expression(paste("S-ML Weibull vs
S-KM")))
> abline(0,1)
>
> plot(Weibml~Utime,type="l",col="red", main=expression(paste("S-
ML Weibull and S-KM vs Time")))
> lines(SKM~Utime,col="black")
```





Παρατηρήσεις:

Με το κόκκινο χρώμα στο παραπάνω γράφημα είναι η κατανομή Weibull ενώ με μαύρο χρώμα είναι οι εκτιμήσεις με την Μέθοδο Μεγίστης Πιθανοφάνειας . Παρατηρούμε λοιπόν ότι η κατανομή Weibull προσαρμόζεται σχετικά καλά στα δεδομένα μου, καθώς οι δύο γραφικές παραστάσεις βρίσκονται πολύ κοντά η μία στην άλλη.

Κατανομή LogNormal.

```
> modlnorm<-fit_data(data = cbdat, dist = "lnorm",    time = "t",
censor = "censor")
> summary(modlnorm)
Fitting of the distribution ' lnorm ' By maximum likelihood on
censored data
Parameters
          estimate Std. Error
meanlog 7.2692490 0.11176661
sdlog   0.4901095 0.08706352
Loglikelihood:  -136.5736   AIC:  277.1472   BIC:  279.1387
Correlation matrix:
            meanlog      sdlog
meanlog 1.00000000 0.08798363
sdlog   0.08798363 1.00000000
```

Συμπέρασμα :

1. **Παράμετροι Κατανομής log-normal:**

Η εκτιμηθείσα τιμή του μέσου (meanlog) είναι περίπου 7.2692490, που αντιπροσωπεύει το μέσο λογαριθμικό της κατανομής log-normal.

Η εκτιμηθείσα τιμή του τυπικού αποκλίνοντος (sdlog) είναι περίπου 0.4901095, που αντιπροσωπεύει το πρότυπο απόκλισης της κατανομής log-normal.

2. **Στατιστικά Κριτήρια:**

Η λογαριθμική πιθανοφάνεια (Loglikelihood) για το μοντέλο είναι περίπου -136.5736.

Το AIC είναι περίπου 277.1472.

Το BIC είναι περίπου 279.1387.

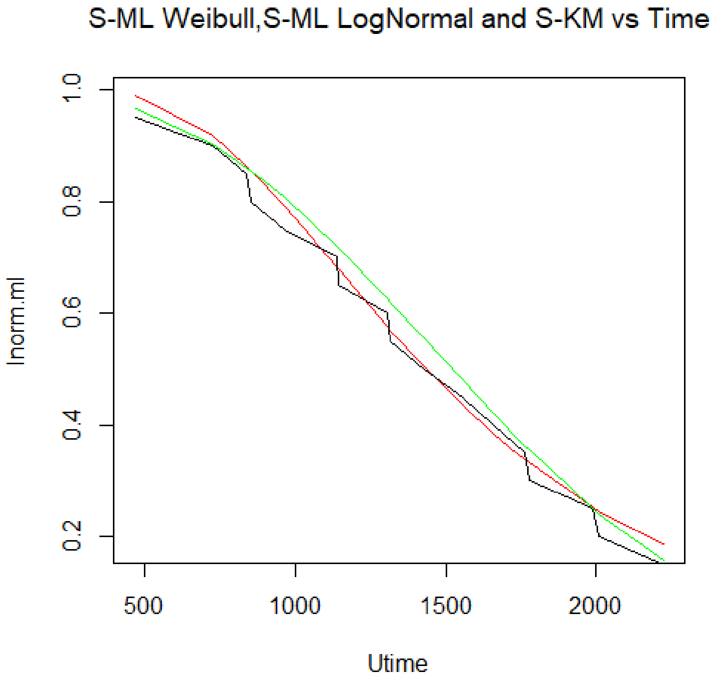
3. **Συσχέτιση Παραμέτρων:**

Η συσχέτιση μεταξύ των παραμέτρων meanlog και sdlog είναι περίπου 0.08798363. Αυτό υποδεικνύει το επίπεδο της συσχέτισης μεταξύ των δύο παραμέτρων.

Συνολικά, τα αποτελέσματα αυτά δείχνουν ότι η κατανομή log-normal προσαρμόζεται καλά στα δεδομένα σας, με τις εκτιμηθείσες παραμέτρους να παρέχουν μια καλή περιγραφή της καμπύλης επιβίωσης.

```
> lnorm.ml<-plnorm(Utime, meanlog=modlnorm$estimate[1] , sdlog =
modlnorm$estimate[2],
+ lower.tail = FALSE, log.p = FALSE)
> lnorm.ml
 [1] 0.9888962 0.9182995 0.8639382 0.8558790 0.7911074 0.6815449
0.6796271
 [8] 0.5776916 0.5697553 0.5048182 0.4357005 0.3843586 0.3370607
0.3320239
[15] 0.2525569 0.2460827 0.1858489
>
>
```

```
> plot(lnorm.ml~Utime,type="l",col="red",
main=expression(paste("S-ML Weibull,S-ML LogNormal and S-KM vs
Time")))
> lines(SKM~Utime,col="black")
> lines(Weibml~Utime,col="green")
```

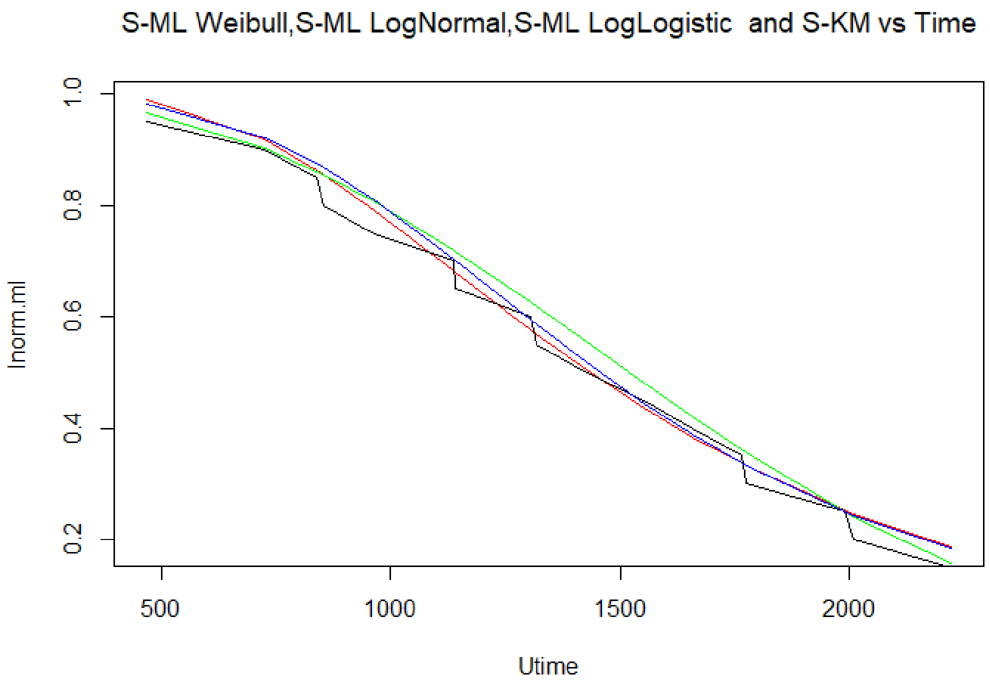


Παρατηρήσεις:

Στο παραπάνω σχήμα τώρα έχουμε στο ίδιο σύστημα αξόνων με μαύρο χρώμα τις εκτιμήσεις από την Μέθοδο Μεγίστης Πιθανοφάνειας , με κόκκινο χρώμα το μοντέλο της κατανομής LogNormal. Και πράσινο χρώμα το μοντέλο της κατανομής Weibull . Σύμφωνα με το σχήμα λοιπόν , το μοντέλο της κατανομής LogNormal προσαρμόζεται πολύ καλά στις εκτιμήσεις μου , ακόμη καλύτερα από το μοντέλο της κατανομής Weibull.

Κατανομή LogLogistic.

```
> library(actuar)
> mod.log.logis<-fit_data(data = cbdat, dist = "llogis",    time
= "t",   censor = "censor")
> summary(mod.log.logis)
Fitting of the distribution ' llogis ' By maximum likelihood on
censored data
Parameters
         estimate Std. Error
shape    3.515627   0.707869
scale 1456.086237 161.513906
Loglikelihood:  -136.6422   AIC:  277.2844   BIC:  279.2759
Correlation matrix:
             shape        scale
shape  1.000000000 -0.001853006
scale -0.001853006  1.000000000

> log.logis.ml <-pllogis(Utime, shape =
mod.log.logis$estimate[1], scale =
+ mod.log.logis$estimate[2], lower.tail = FALSE, log.p = FALSE)
> log.logis.ml
 [1] 0.9818430 0.9206764 0.8746097 0.8676082 0.8094216 0.7033796
0.7014465
 [8] 0.5957599 0.5873337 0.5177270 0.4430489 0.3878072 0.3375103
0.3322017
[15] 0.2500735 0.2435385 0.1840625

> plot(lnorm.ml~Utime,type="l", col="red",
main=expression(paste("S-ML Weibull,S-ML LogNormal,S-ML
LogLogistic  and S-KM vs Time")))
> lines(SKM~Utime,col="black")
> lines(Weibml~Utime,col="green")
> lines(log.logis.ml~Utime,col="blue")
```



Παρατηρήσεις:

Στο παραπάνω σχήμα τώρα έχουμε στο ίδιο σύστημα αξόνων με μαύρο χρώμα τις εκτιμήσεις από την Μέθοδο Μεγίστης Πιθανοφάνειας , με κόκκινο χρώμα το μοντέλο της κατανομής LogNormal, με πράσινο χρώμα το μοντέλο της κατανομής Weibull και με μπλε χρώμα το μοντέλο της κατανομής LogLogistic Σύμφωνα με το σχήμα λοιπόν , το μοντέλο της κατανομής LogLogistic προσαρμόζεται πολύ καλά στις εκτιμήσεις μου , ακόμη καλύτερα από το μοντέλο της κατανομής Weibull. Ωστοσο το μοντέλο της κατανομής LogNormal φαίνεται να είναι το καλύτερπ .

iv. **Να γίνει ο έλεγχος Anderson–Darling για κάθε μια από αυτές τις κατανομές και να εφαρμοστεί το κριτήριο AIC.**

Σε αυτόν τον έλεγχο θα συγκρίνω τα αποτελέσματα του Anderson-Darling Test και του AIC Test για τις κατανομες Weibull, Log Normal και Log Logistic.
Δεν θα κάνω κάποια σύγκριση με την Exponential , καθώς τα αποτελέσματα της για την προσαρμογή της με parmsurvfit δεν είναι πολύ καλά . Άλλωστε στα προηγούμενα ερωτήματα απέρριψα το μοντέλο της, δηλαδή τα δεδομένα μου αποκλείεται να ακολουθούν το μοντέλο της Εκθετικής.

```
> compute_AD (data = cbdat, dist = "weibull",    time = "t",
censor = "censor")
[1] 13.78981
> compute_AD (data = cbdat, dist = "lnorm",    time = "t",
censor = "censor")
[1] 13.83252
> compute_AD (data = cbdat, dist = "llogis",    time = "t",
censor = "censor")
[1] 13.80438
```

	AD	AIC
Weibull	13.78981	277.2924
LogNormal	13.83252	277.1472
LogLogistic	13.80438	277.2844

Παρατηρήσεις:

Μια μικρή τιμή του AD τεστ υποδηλώνει ότι τα δεδομένα σας πιθανότατα προέρχονται από την κατανομή που ελέγχεται, ενώ μια μεγαλύτερη τιμή υποδηλώνει μεγαλύτερες αποκλίσεις από αυτήν την κατανομή.

Με βάση τις τιμές που παρέχονται, φαίνεται ότι η Weibull κατανομή έχει το χαμηλότερο Anderson-Darling τεστ, οπότε θα μπορούσε να είναι η πλέον κατάλληλη για τα δεδομένα σας. Ωστόσο, οι αποκλίσεις από τη Weibull κατανομή είναι ελάχιστες μεταξύ των τριών επιλογών, οπότε ενδέχεται να απαιτούνται περαιτέρω εξέταση.

Το κριτήριο του Πληροφοριακού Κριτηρίου (AIC) χρησιμοποιείται για τη σύγκριση των μοντέλων, με μικρότερες τιμές να υποδεικνύουν καλύτερη προσαρμογή των δεδομένων. Από τις παρεχόμενες τιμές AIC, φαίνεται ότι η Log-Normal κατανομή έχει τη μικρότερη τιμή AIC (277.1472), ενώ η Weibull και η Log-Logistic έχουν τιμές AIC 277.2924 και 277.2844 αντίστοιχα.Ωστόσο, η διαφορά μεταξύ των τιμών AIC είναι μικρή, οπότε πρέπει να ληφθούν υπόψη και άλλοι παράγοντες κατά τη λήψη αποφάσεων για την κατάλληλη κατανομή.

v. **Να γίνει ο στατιστικός έλεγχος υποθέσεων της Εκθετικής κατανομής (η = 1) με εναλλακτική την κατανομή Weibull (η ≠ 1) (μπορείτε να αξιοποιήσετε και ένα 0.95 Δ.Ε. για την παράμετρο η της κατανομής Weibull). Τι συμπεραίνετε;**

Wald Test

Κάνει ένα τεστ και ελέγχει αν ο λογάριθμος του α είναι 0 ή όχι, δηλαδή κάνει ένα τεστ για να δει εάν τα δεδομένα μας προσαρμόζονται καλύτερα στο μοντέλο της Εκθετικης Κατανομής ή στο μοντέλο της Κατανομής Weibull.

```
> summary(mod1)

Call:
survreg(formula = Surv(t, censor) ~ 1, data = cbdat, dist =
"exponential")
            Value Std. Error    z      p
(Intercept) 7.475      0.243 30.8 <2e-16

Scale fixed at 1

Exponential distribution
Loglik(model)= -144.1   Loglik(intercept only)= -144.1
Number of Newton-Raphson Iterations: 3
n= 20
```

<u>Συμπέρασμα :</u>

Το p-value είναι πολύ μικρό , άρα απορρίπτω την μηδενική υπόθεση. Άρα συμπεραίνω ότι τα δεδομένα ακολουθούν καλύτερα το μοντέλο της Weibull και όχι της Εκθετικης.

(C) Τα δεδομένα στο αρχείο «lung-patients.txt» αφορούν τη διάρκεια παροχής οξυγόνου (σε ημέρες) σε 78 ασθενείς με αναπνευστικά προβλήματα που νοσηλεύτηκαν σε κλινική για ένα συγκεκριμένο διάστημα. Η μία ομάδα των 43 ασθενών υποβλήθηκε σε ενισχυμένη θεραπεία πνεύμονα (=0), ενώ οι υπόλοιποι 35 όχι (=1).

Ξεκινάω με εισαγωγή των δεδομένων στην R .

```
> library(splines)
> library(survival)
> cc <- read.table("C:\\Users\\Admin\\Desktop\\8o ΕΞΑΜΗΜΟ\\lung-patients.txt", header = TRUE)
> attach(cc)
The following objects are masked from cbdat (pos = 4):

    c, Group, id, t

The following object is masked from cbdat (pos = 5):

    t

The following object is masked from cbdat (pos = 7):

    t

The following object is masked from cbdat (pos = 8):

    t

> cc
   id Group   t c
1   1     0  59 1
2   2     0 514 1
3   3     0 313 1
4   4     0 631 1
5   5     0 107 1
6   6     0  71 1
7   7     0 583 1
8   8     0  91 1
9   9     0  66 1
10 10     0  95 1
11 11     0  13 1
12 12     0   5 1
13 13     0  85 1
14 14     0 619 0
15 15     0 580 1
16 16     0 196 1
17 17     0 475 1
18 18     0  32 1
19 19     0 161 1
20 20     0 193 0
21 21     0  59 1
22 22     0  62 1
23 23     0  95 1
24 24     0  63 1
25 25     0  26 1
```

```
26 26      0  16 1
27 27      0 553 1
28 28      0  76 1
29 29      0 134 1
30 30      0 116 1
31 31      0  83 1
32 32      0  33 1
33 33      0 317 1
34 34      0 600 1
35 35      0 362 1
36 36      0 333 1
37 37      0  68 1
38 38      0 217 1
39 39      0 733 0
40 40      0 546 1
41 41      0 546 1
42 42      0  56 1
43 43      0  48 1
44 44      1  43 1
45 45      1 250 1
46 46      1 110 1
47 47      1 249 1
48 48      1 181 1
49 49      1  70 1
50 50      1 197 1
51 51      1 306 1
52 52      1  53 1
53 53      1  30 1
54 54      1  45 1
55 55      1  23 1
56 56      1  54 1
57 57      1  63 1
58 58      1  14 1
59 59      1  96 1
60 60      1 103 1
61 61      1  71 1
62 62      1  71 1
63 63      1  64 1
64 64      1 253 1
65 65      1  54 1
66 66      1 236 1
67 67      1  51 1
68 68      1 134 1
69 69      1  31 1
70 70      1 274 0
71 71      1 204 1
72 72      1 118 1
73 73      1 424 1
74 74      1  56 1
75 75      1 310 0
76 76      1 108 1
77 77      1  51 1
78 78      1  70 1
>
```

i. Να υπολογιστούν οι εκτιμήτριες Kaplan–Meier και να γίνουν οι γραφικές παραστάσεις αυτών για τις δύο ομάδες.

Στο ερώτημα αυτό θα υπολογίσω τις εκτιμήσεις μέσω της μη παραμετρικη μεθόδου Kaplan Meier .

```
> outp<-survfit(Surv(t,c)~Group)
> summary(outp)
Call: survfit(formula = Surv(t, c) ~ Group)

                Group=0
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    5     43       1   0.9767  0.0230      0.93272        1.000
   13     42       1   0.9535  0.0321      0.89258        1.000
   16     41       1   0.9302  0.0388      0.85712        1.000
   26     40       1   0.9070  0.0443      0.82418        0.998
   32     39       1   0.8837  0.0489      0.79292        0.985
   33     38       1   0.8605  0.0528      0.76289        0.971
   48     37       1   0.8372  0.0563      0.73383        0.955
   56     36       1   0.8140  0.0593      0.70557        0.939
   59     35       2   0.7674  0.0644      0.65101        0.905
   62     33       1   0.7442  0.0665      0.62456        0.887
   63     32       1   0.7209  0.0684      0.59859        0.868
   66     31       1   0.6977  0.0700      0.57306        0.849
   68     30       1   0.6744  0.0715      0.54795        0.830
   71     29       1   0.6512  0.0727      0.52322        0.810
   76     28       1   0.6279  0.0737      0.49885        0.790
   83     27       1   0.6047  0.0746      0.47483        0.770
   85     26       1   0.5814  0.0752      0.45116        0.749
   91     25       1   0.5581  0.0757      0.42781        0.728
   95     24       2   0.5116  0.0762      0.38206        0.685
  107     22       1   0.4884  0.0762      0.35966        0.663
  116     21       1   0.4651  0.0761      0.33757        0.641
  134     20       1   0.4419  0.0757      0.31579        0.618
  161     19       1   0.4186  0.0752      0.29432        0.595
  196     17       1   0.3940  0.0747      0.27166        0.571
  217     16       1   0.3694  0.0740      0.24940        0.547
  313     15       1   0.3447  0.0731      0.22757        0.522
  317     14       1   0.3201  0.0719      0.20616        0.497
  333     13       1   0.2955  0.0704      0.18521        0.471
  362     12       1   0.2709  0.0687      0.16473        0.445
  475     11       1   0.2462  0.0667      0.14475        0.419
  514     10       1   0.2216  0.0645      0.12532        0.392
  546      9       2   0.1724  0.0588      0.08833        0.336
  553      7       1   0.1477  0.0553      0.07093        0.308
  580      6       1   0.1231  0.0513      0.05442        0.279
  583      5       1   0.0985  0.0466      0.03900        0.249
  600      4       1   0.0739  0.0409      0.02495        0.219
  631      2       1   0.0369  0.0332      0.00635        0.215
```

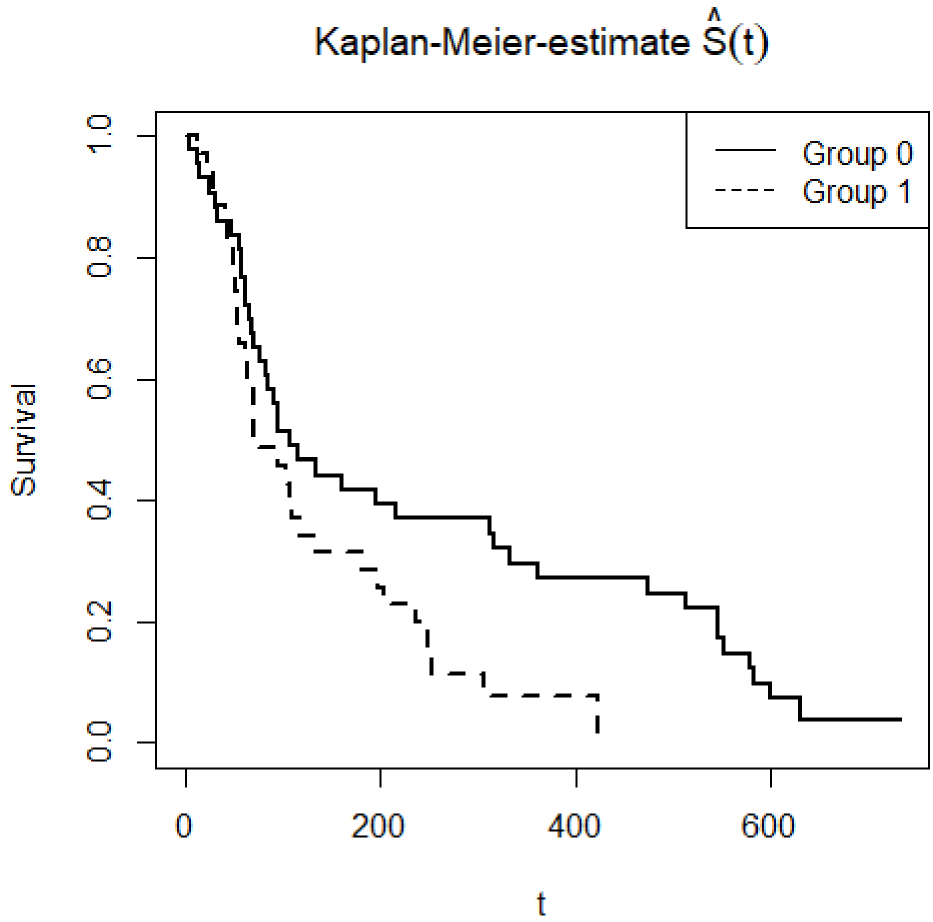
```
              Group=1
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   14     35       1   0.9714  0.0282       0.9178        1.000
   23     34       1   0.9429  0.0392       0.8690        1.000
   30     33       1   0.9143  0.0473       0.8261        1.000
   31     32       1   0.8857  0.0538       0.7863        0.998
   43     31       1   0.8571  0.0591       0.7487        0.981
   45     30       1   0.8286  0.0637       0.7127        0.963
   51     29       2   0.7714  0.0710       0.6441        0.924
   53     27       1   0.7429  0.0739       0.6113        0.903
   54     26       2   0.6857  0.0785       0.5479        0.858
   56     24       1   0.6571  0.0802       0.5173        0.835
   63     23       1   0.6286  0.0817       0.4873        0.811
   64     22       1   0.6000  0.0828       0.4578        0.786
   70     21       2   0.5429  0.0842       0.4005        0.736
   71     19       2   0.4857  0.0845       0.3454        0.683
   96     17       1   0.4571  0.0842       0.3186        0.656
  103     16       1   0.4286  0.0836       0.2923        0.628
  108     15       1   0.4000  0.0828       0.2666        0.600
  110     14       1   0.3714  0.0817       0.2414        0.572
  118     13       1   0.3429  0.0802       0.2167        0.542
  134     12       1   0.3143  0.0785       0.1927        0.513
  181     11       1   0.2857  0.0764       0.1692        0.482
  197     10       1   0.2571  0.0739       0.1464        0.452
  204      9       1   0.2286  0.0710       0.1244        0.420
  236      8       1   0.2000  0.0676       0.1031        0.388
  249      7       1   0.1714  0.0637       0.0827        0.355
  250      6       1   0.1429  0.0591       0.0635        0.322
  253      5       1   0.1143  0.0538       0.0454        0.287
  306      3       1   0.0762  0.0475       0.0225        0.258
  424      1       1   0.0000     NaN           NA           NA
```

Παρατηρήσεις:

Είναι φανερό ότι όσο περνάει ο χρόνος όλο και λιγότεροι επιβιώνουν .

Ύστερα θα κατασκευάσω την γραφική παράσταση των εκτιμήσεων της Kaplan – Meier .

```
> plot(outp, lty = 1:2, main=expression(paste("Kaplan-Meier-estimate ",
hat(S)(t))),
+ xlab="t", ylab="Survival", lwd=2)
> legend("topright", c("Group 0", "Group 1"), lty = 1:2)
>
```



Παρατηρήσεις:

Φαίνεται σύμφωνα με το παραπάνω γράφημα ότι οι ασθένεις από την πρώτη ομάδα (ομάδα =0) των ασθένων , δηλαδη από την ομάδα που υποβληθηκε σε ενισχυμένη θεραπεία , επιβιώνουν περισσότερο , από τους ασθενεις που βρίσκονται στην δεύτερη ομάδα (ομάδα =1), οι οποίοι δεν υποβληθηκαν σε ενισχυμένη θεραπεία.

ii. **Να γίνουν οι έλεγχοι log–rank και Wilcoxon για τη σύγκριση των δύο ομάδων ως προς τη διάρκεια παροχής οξυγόνου.**

```
> #Log-rank test (Mantel-Haenszel test) with R
> out1<-survdiff(Surv(t, c) ~ Group)
> out1
Call:
survdiff(formula = Surv(t, c) ~ Group)

         N Observed Expected (O-E)^2/E (O-E)^2/V
Group=0 43       40     48.9      1.64      5.62
Group=1 35       33     24.1      3.33      5.62

 Chisq= 5.6  on 1 degrees of freedom, p= 0.02
>
```

Παρατηρήσεις:

Η συνολική στατιστική τιμή χι-τετραγωνικού είναι 5.6 με 1 βαθμό ελευθερίας και η p-τιμή είναι 0.02. Αυτό υποδεικνύει ότι υπάρχει στατιστικά σημαντική διαφορά μεταξύ των επιβιώσεων των δύο ομάδων στο επίπεδο εμπιστοσύνης του 95%, καθώς η p-τιμή είναι μικρότερη από το επιπεδο αποκοπής 0.05.

iii. **Κάνοντας χρήση των εκτιμήσεων Kaplan–Meier να γίνουν γραφικοί έλεγχοι των κατανομών Weibull και Λογαριθμο–λογιστικής για τις δύο ομάδες χωριστά. Ενισχύστε τα συμπεράσματά σας με το κριτήριο AIC.**

https://helios.ntua.gr/pluginfile.php/264283/mod_resource/content/3/Hodgins1-%20parametric%20results-weibull%20and%20loglogistic.pdf

```
> library(survival)
> library(parmsurvfit)
```

Parametric Results for 1st group

```
> group0<- Surv(t[Group=="0"],c[Group=="0"])
> outp0<-survfit(group0~1, type="kaplan-meier",data=cc)
> summary(outp0)
Call: survfit(formula = group0 ~ 1, data = cc, type = "kaplan-meier")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    5     43       1   0.9767  0.0230      0.93272        1.000
   13     42       1   0.9535  0.0321      0.89258        1.000
   16     41       1   0.9302  0.0388      0.85712        1.000
   26     40       1   0.9070  0.0443      0.82418        0.998
   32     39       1   0.8837  0.0489      0.79292        0.985
   33     38       1   0.8605  0.0528      0.76289        0.971
   48     37       1   0.8372  0.0563      0.73383        0.955
   56     36       1   0.8140  0.0593      0.70557        0.939
   59     35       2   0.7674  0.0644      0.65101        0.905
   62     33       1   0.7442  0.0665      0.62456        0.887
   63     32       1   0.7209  0.0684      0.59859        0.868
   66     31       1   0.6977  0.0700      0.57306        0.849
   68     30       1   0.6744  0.0715      0.54795        0.830
   71     29       1   0.6512  0.0727      0.52322        0.810
   76     28       1   0.6279  0.0737      0.49885        0.790
   83     27       1   0.6047  0.0746      0.47483        0.770
   85     26       1   0.5814  0.0752      0.45116        0.749
   91     25       1   0.5581  0.0757      0.42781        0.728
   95     24       2   0.5116  0.0762      0.38206        0.685
  107     22       1   0.4884  0.0762      0.35966        0.663
  116     21       1   0.4651  0.0761      0.33757        0.641
  134     20       1   0.4419  0.0757      0.31579        0.618
  161     19       1   0.4186  0.0752      0.29432        0.595
  196     17       1   0.3940  0.0747      0.27166        0.571
  217     16       1   0.3694  0.0740      0.24940        0.547
  313     15       1   0.3447  0.0731      0.22757        0.522
  317     14       1   0.3201  0.0719      0.20616        0.497
  333     13       1   0.2955  0.0704      0.18521        0.471
  362     12       1   0.2709  0.0687      0.16473        0.445
  475     11       1   0.2462  0.0667      0.14475        0.419
  514     10       1   0.2216  0.0645      0.12532        0.392
  546      9       2   0.1724  0.0588      0.08833        0.336
  553      7       1   0.1477  0.0553      0.07093        0.308
  580      6       1   0.1231  0.0513      0.05442        0.279
  583      5       1   0.0985  0.0466      0.03900        0.249
  600      4       1   0.0739  0.0409      0.02495        0.219
  631      2       1   0.0369  0.0332      0.00635        0.215
```

Split data into two groups

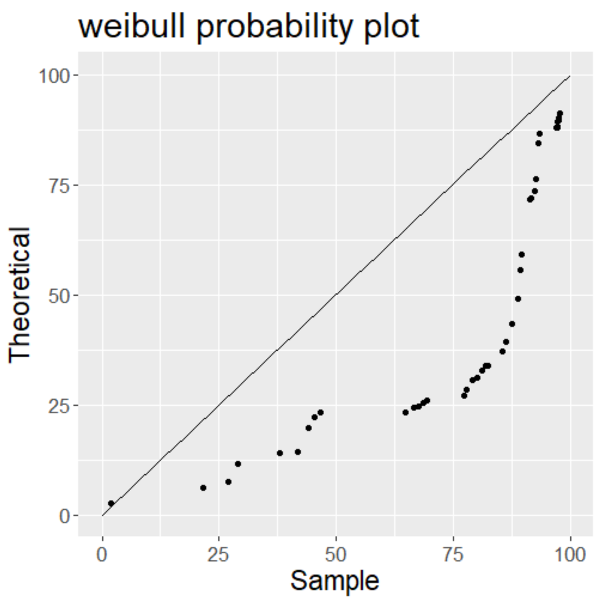
```
> ccdat0 <- subset(cc, Group ==0)
> ccdat1 <-subset(cc, Group ==1)
```

Fitting Weibull distribution

```
> modweib0<-fit_data(data = ccdat0, dist = "weibull", time = "t", censor
= "c")
> summary(modweib0)
Fitting of the distribution ' weibull ' By maximum likelihood on censored
data
Parameters
         estimate Std. Error
shape   0.9338227  0.1179091
scale 244.8090140 42.8450786
Loglikelihood:  -260.8308   AIC:  525.6616   BIC:  529.184
Correlation matrix:
          shape     scale
shape 1.0000000 0.2515068
scale 0.2515068 1.0000000

> SWeib.ml0<- pweibull(Utime0, shape= modweib0$estimate[1], scale =
+ modweib0$estimate[2], lower.tail = FALSE, log.p = FALSE)
> SWeib.ml0
 [1] 0.97392358 0.93754661 0.92469838 0.88409127 0.86109077 0.85734471
 [7] 0.80380897 0.77708419 0.75778601 0.75462870 0.74525487 0.73908593
[13] 0.72995014 0.71502875 0.69475369 0.68908785 0.67241664 0.63022242
[19] 0.60783752 0.56573128 0.50857139 0.44375052 0.40921817 0.28424327
[25] 0.28000988 0.26373012 0.23671066 0.15613762 0.13546767 0.11761777
[31] 0.10669888 0.10555202 0.09928900 0.08883526

> plot_ppsurv(ccdat0, "weibull", time = "t",  censor = "c")
```

Συμπέρασμα :

Σύμφωνα με το διαγγραμμα στα δεξία, λόγω δεν παρατηρώ κάποια τάυτιση των αποτελεσματων (εκτιμησεων KM) με την ευθεία έχω αρκετές ενδείξης ώστε να απορριψω την υποθέση , ότι τα δεδομενα της πρώτης ομάδας ακολουθούν το μοντέλο της Weibull Κατανομή.

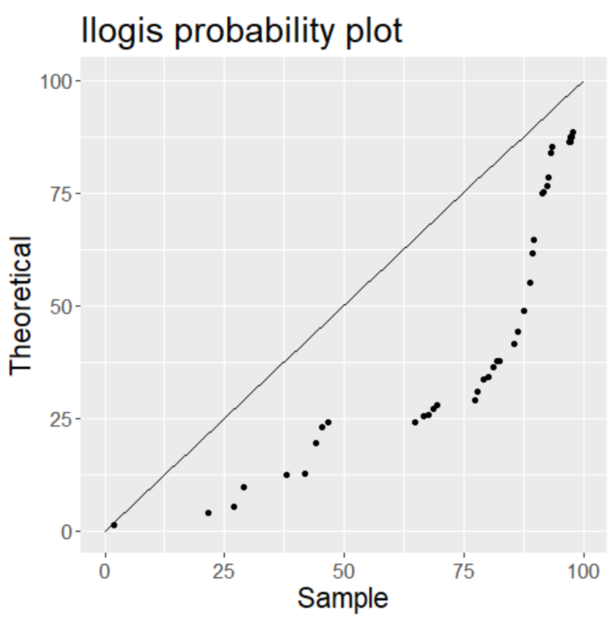


Fitting Log-logistic distribution

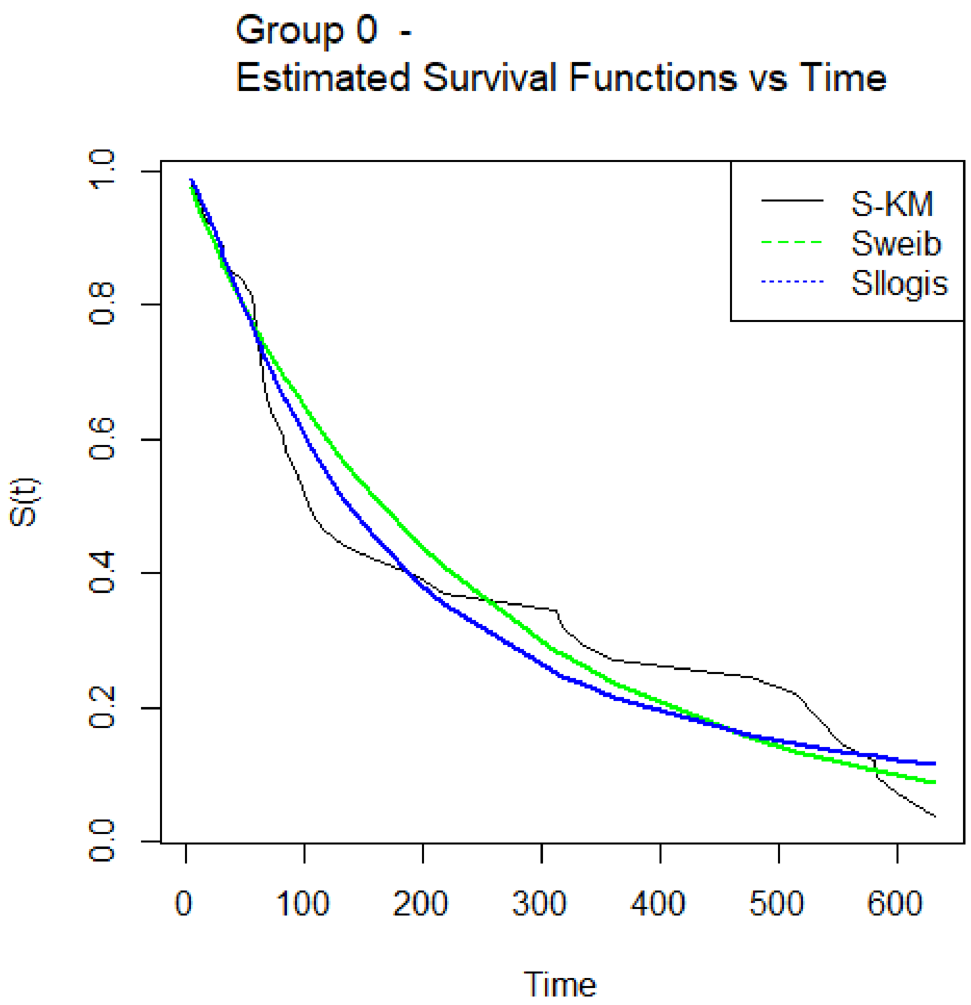
```
> library(actuar)
>
> modllog0<-fit_data(data = ccdat0, dist = "llogis", time = "t", censor =
"c")
> summary(modllog0)
Fitting of the distribution ' llogis ' By maximum likelihood on censored
data
Parameters
        estimate Std. Error
shape   1.341917  0.1729928
scale 138.446997 27.8095493
Loglikelihood:  -261.1803   AIC:  526.3605   BIC:  529.8829
Correlation matrix:
            shape       scale
shape  1.00000000 -0.02933534
scale -0.02933534  1.00000000

> Sllogis.ml0<-pllogis(Utime0, shape = modllog0$estimate[1], scale =
+ modllog0$estimate[2], lower.tail = FALSE, log.p = FALSE)
> Sllogis.ml0
[1]  0.9885310 0.9598578 0.9476345 0.9041500 0.8771351 0.8726152 0.8055695
[8]  0.7711149 0.7461217 0.7420331 0.7299042 0.7219342 0.7101562 0.6910047
[15] 0.6652075 0.6580544 0.6371709 0.5855882 0.5590683 0.5109509
0.4495429
[22] 0.3854481 0.3536415 0.2507488 0.2475609 0.2354587 0.2158879
0.1605216
[29] 0.1467602 0.1348918 0.1275996 0.1268310 0.1226207 0.1155314

> plot_ppsurv(ccdat0, "llogis", time = "t",  censor = "c")
```

Συμπέρασμα :

Σύμφωνα με το διαγγραμμα στα δεξία, λόγω δεν παρατηρώ κάποια τάυτιση των αποτελεσματων (εκτιμησεων KM) με την ευθεία έχω αρκετές ενδείξης ώστε να απορρίψω την υποθέση , ότι τα δεδομένα της πρώτης ομάδας ακολουθούν το μοντέλο της **Log-logistic** Κατανομή.



```
> plot(SKM0~ Utime0,type="l", col="black", main=expression(paste("Group 0
-
+ Estimated Survival Functions vs Time")), xlab="Time", ylab="S(t)")
> lines(SWeib.ml0~ Utime0,col="green", lwd=2)
> lines(Sllogis.ml0~ Utime0,col="blue", lwd=2)
> legend("topright", c("S-KM", "Sweib", "Sllogis"), col=c("black",
"green","blue"), lty=c(1,2,3) )
```



Συμπέρασμα :

Στο παραπάνω σχήμα η μαύρη καμπύλη παρουσιάζει την εκτίμηση μέσω της Kaplan – Meier , η πράσινη καμπύλη παρουσιάζει το μοντέλο της κατανομής Weibull και η μπλέ καμπύλη παρουσιάζει το μοντέλο της κατανομής Log-logistic για τα δεδομένα της 1ης ομάδας (Group = 0).

Είναι φανερό στο παραπάνω σχήμα ότι , κανένα από τα δύο μοντέλα κακόνομων δεν προσαρμόζεται καλά στα δεδομένα της 1ης ομάδας (Group = 0).

ΚΡΙΤΗΡΙΟ AIC

1st Group (Group = 0)	
	AIC
Weibull Distribution	525.6616
Log-Logistic Distribution	526.3605

Συμπέρασμα :

Βασισμένο στα παραπάνω αποτελέσματα, φαίνεται ότι το μοντέλο Log-Logistic Distribution έχει μικρότερο AIC (526.3605) σε σύγκριση με το μοντέλο Weibull Distribution (525.6616) για την 1η ομάδα (Group = 0). Συνεπώς, το μοντέλο Log-Logistic φαίνεται να προσαρμόζεται καλύτερα στα δεδομένα της 1ης ομάδας, βάσει του κριτηρίου του AIC.

Ωστόσο το AIC είναι πολύ υψηλό και στις δύο περιπτώσεις , άρα συμπεραίνω ότι κανένα μοντέλο δεν είναι το ιδανικό για τα δεδομένα της 1ης ομάδας (Group = 0).

Parametric Results for 2nd group

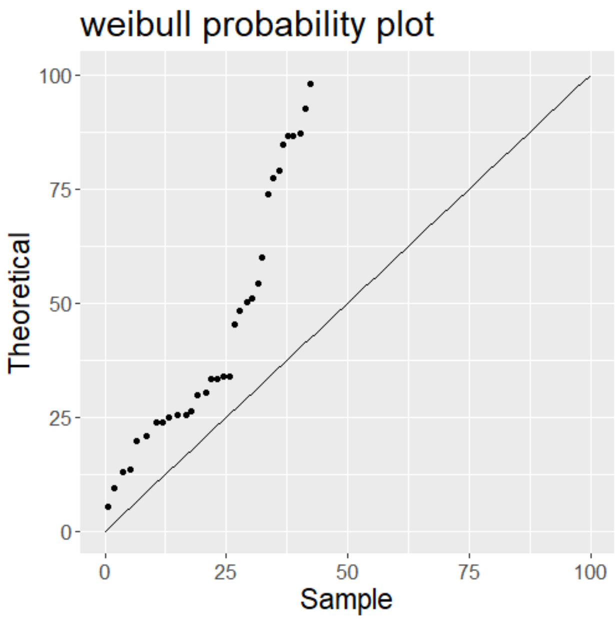
```
> group1<- Surv(t[Group=="1"],c[Group=="1"])
> outp1<-survfit(group1~1, type="kaplan-meier",data=cc)
> summary(outp1)
Call: survfit(formula = group1 ~ 1, data = cc, type = "kaplan-meier")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   14     35       1   0.9714  0.0282       0.9178        1.000
   23     34       1   0.9429  0.0392       0.8690        1.000
   30     33       1   0.9143  0.0473       0.8261        1.000
   31     32       1   0.8857  0.0538       0.7863        0.998
   43     31       1   0.8571  0.0591       0.7487        0.981
   45     30       1   0.8286  0.0637       0.7127        0.963
   51     29       2   0.7714  0.0710       0.6441        0.924
   53     27       1   0.7429  0.0739       0.6113        0.903
   54     26       2   0.6857  0.0785       0.5479        0.858
   56     24       1   0.6571  0.0802       0.5173        0.835
   63     23       1   0.6286  0.0817       0.4873        0.811
   64     22       1   0.6000  0.0828       0.4578        0.786
   70     21       2   0.5429  0.0842       0.4005        0.736
   71     19       2   0.4857  0.0845       0.3454        0.683
   96     17       1   0.4571  0.0842       0.3186        0.656
  103     16       1   0.4286  0.0836       0.2923        0.628
  108     15       1   0.4000  0.0828       0.2666        0.600
  110     14       1   0.3714  0.0817       0.2414        0.572
  118     13       1   0.3429  0.0802       0.2167        0.542
  134     12       1   0.3143  0.0785       0.1927        0.513
  181     11       1   0.2857  0.0764       0.1692        0.482
  197     10       1   0.2571  0.0739       0.1464        0.452
  204      9       1   0.2286  0.0710       0.1244        0.420
  236      8       1   0.2000  0.0676       0.1031        0.388
  249      7       1   0.1714  0.0637       0.0827        0.355
  250      6       1   0.1429  0.0591       0.0635        0.322
  253      5       1   0.1143  0.0538       0.0454        0.287
  306      3       1   0.0762  0.0475       0.0225        0.258
  424      1       1   0.0000     NaN           NA           NA
>

> Utime1<- outp1$time[outp1$n.event==1]
> SKM1<-outp1$surv[outp1$n.event==1]
>
```

Fitting Weibull distribution

```
> modweib1<-fit_data(data = ccdat1, dist = "weibull", time = "t", censor
="c")
> summary(modweib1)
Fitting of the distribution ' weibull ' By maximum likelihood on censored
data
Parameters
        estimate Std. Error
shape   1.259085  0.1699116
scale 143.685225 20.6333246
Loglikelihood:  -193.6603   AIC:  391.3207   BIC:  394.4314
Correlation matrix:
          shape     scale
shape 1.0000000 0.2698941
scale 0.2698941 1.0000000

> SWeib.ml1<- pweibull(Utime1, shape= modweib1$estimate[1], scale =
+ modweib1$estimate[2], lower.tail = FALSE, log.p = FALSE)
> SWeib.ml1
 [1] 0.94809789 0.90521889 0.87010516 0.86501719 0.80337595 0.79308092
 [7] 0.75211312 0.73688925 0.70178681 0.69682744 0.54780151 0.51808851
[13] 0.49755336 0.48950166 0.45822842 0.40016270 0.26254133 0.22585488
[19] 0.21124600 0.15446152 0.13557003 0.13420637 0.13018898 0.07498887
[25] 0.02012451

> plot_ppsurv(ccdat1, "weibull", time = "t",  censor = "c")
>
```

Συμπέρασμα :

Σύμφωνα με το διαγγραμμα στα δεξία, λόγω δεν παρατηρώ κάποια τάυτιση των αποτελεσματων (εκτιμησεων KM) με την ευθεία έχω αρκετές ενδείξης ώστε να απορρίψω την υποθέση , ότι τα δεδομένα της δεύτερης ομάδας ακολουθούν το μοντέλο της **Weibull** Κατανομή.

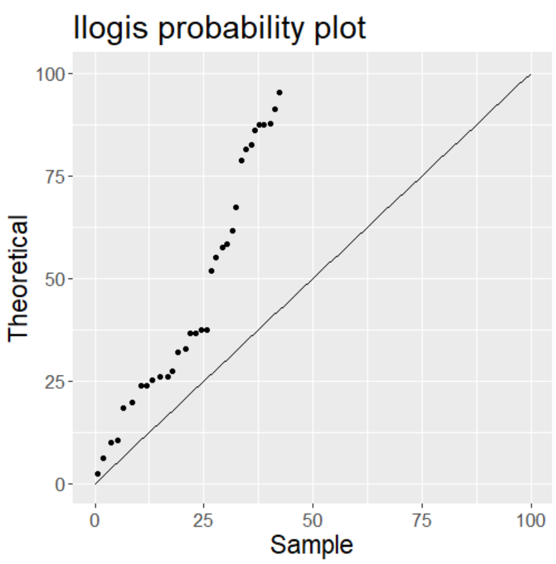


Fitting Log-logistic distribution

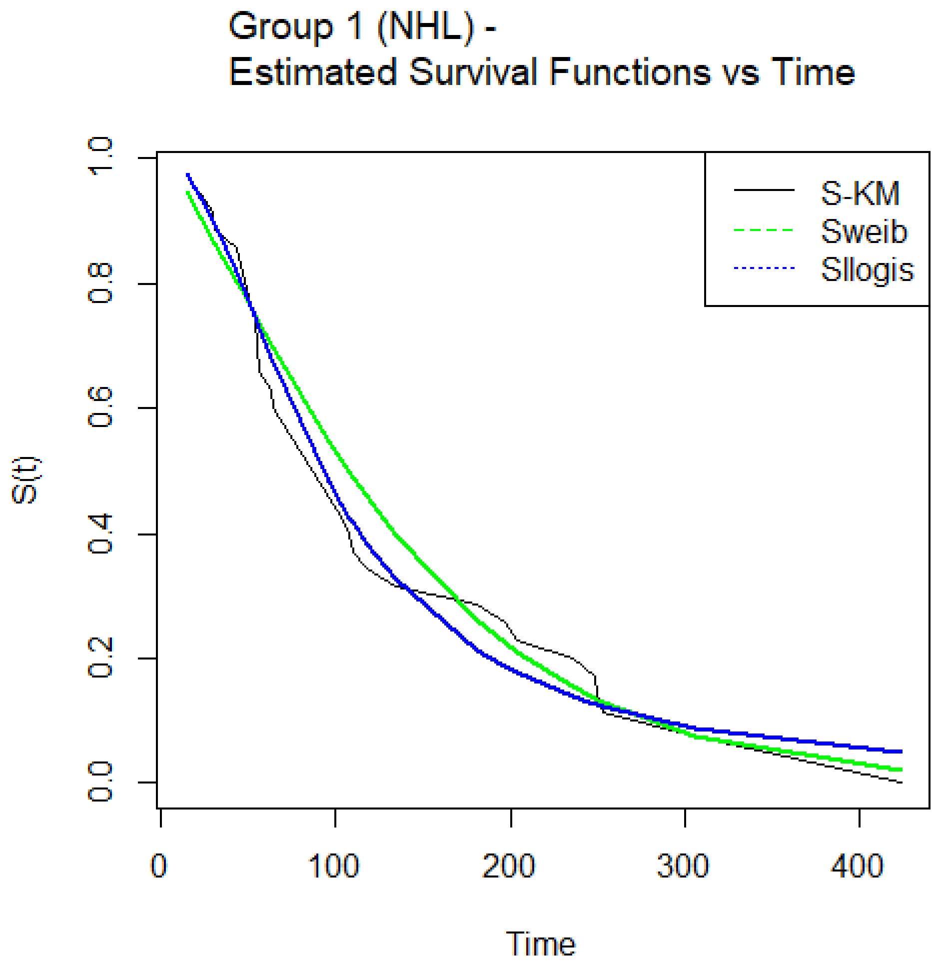
```
> library(actuar)

Attaching package: 'actuar'

The following objects are masked from 'package:stats':

    sd, var

The following object is masked from 'package:grDevices':

    cm

> modllog1<-fit_data(data = ccdat1, dist = "llogis", time = "t", censor = "c")
> summary(modllog1)
Fitting of the distribution ' llogis ' By maximum likelihood on censored data
Parameters
       estimate Std. Error
shape  1.953817  0.2782326
scale 92.627407 14.1293125
Loglikelihood:  -192.2891   AIC:  388.5781   BIC:  391.6888
Correlation matrix:
            shape       scale
shape  1.00000000 -0.05036393
scale -0.05036393  1.00000000


> Sllogis.ml1<-pllogis(Utime1, shape = modllog1$estimate[1], scale =
+ modllog1$estimate[2], lower.tail = FALSE, log.p = FALSE)
> Sllogis.ml1
 [1] 0.97567897 0.93830343 0.90049257 0.89460285 0.81747453 0.80384556
 [7] 0.74853287 0.72774788 0.67985370 0.67311993 0.48253845 0.44833857
[13] 0.42555721 0.41681743 0.38390210 0.32707238 0.21267338 0.18627673
[19] 0.17615661 0.13856008 0.12652196 0.12565906 0.12312069 0.08828061
[25] 0.04870484

> plot_ppsurv(ccdat1, "llogis", time = "t",  censor = "c")
>
```

Συμπέρασμα :

Σύμφωνα με το διαγραμμα στα δεξία, λόγω δεν παρατηρώ κάποια τάυτιση των αποτελεσματων (εκτιμησεων ΚΜ) με την ευθεία έχω αρκετές ενδείξης ώστε να απορρίψω την υποθέση , ότι τα δεδομένα της δεύτερής ομάδας ακολουθούν το μοντέλο της **Log-logistic** Κατανομή.



```
> plot(SKM1~ Utime1,type="l", col="black", main=expression(paste("Group 1
(NHL) -
+ Estimated Survival Functions vs Time")), xlab="Time", ylab="S(t)")
> lines(SWeib.ml1~ Utime1,col="green", lwd=2)
> lines(Sllogis.ml1~ Utime1,col="blue", lwd=2)
> legend("topright", c("S-KM", "Sweib", "Sllogis"), col=c("black",
"green","blue"), lty=c(1,2,3) )
>
```



Συμπέρασμα :

Στο παραπάνω σχήμα η μαύρη καμπύλη παρουσιάζει την εκτίμηση μέσω της Kaplan – Meier , η πράσινη καμπύλη παρουσιάζει το μοντέλο της κατανομής Weibull και η μπλέ καμπύλη παρουσιάζει το μοντέλο της κατανομής Log-logistic για τα δεδομένα της 2ης ομάδας (Group = 1).

Είναι φανερό στο παραπάνω σχήμα ότι , το μοντέλο της κατανομής Log-logistic προσαρμόζεται πολύ καλά στα δεδομένα της 1ης ομάδας (Group = 0), καλύτερα από το μοντέλο της κατανομής Weibull .

ΚΡΙΤΗΡΙΟ AIC

2nd Group (Group = 1)	
	AIC
Weibull Distribution	391.3207
Log-Logistic Distribution	388.5781

Συμπέρασμα :

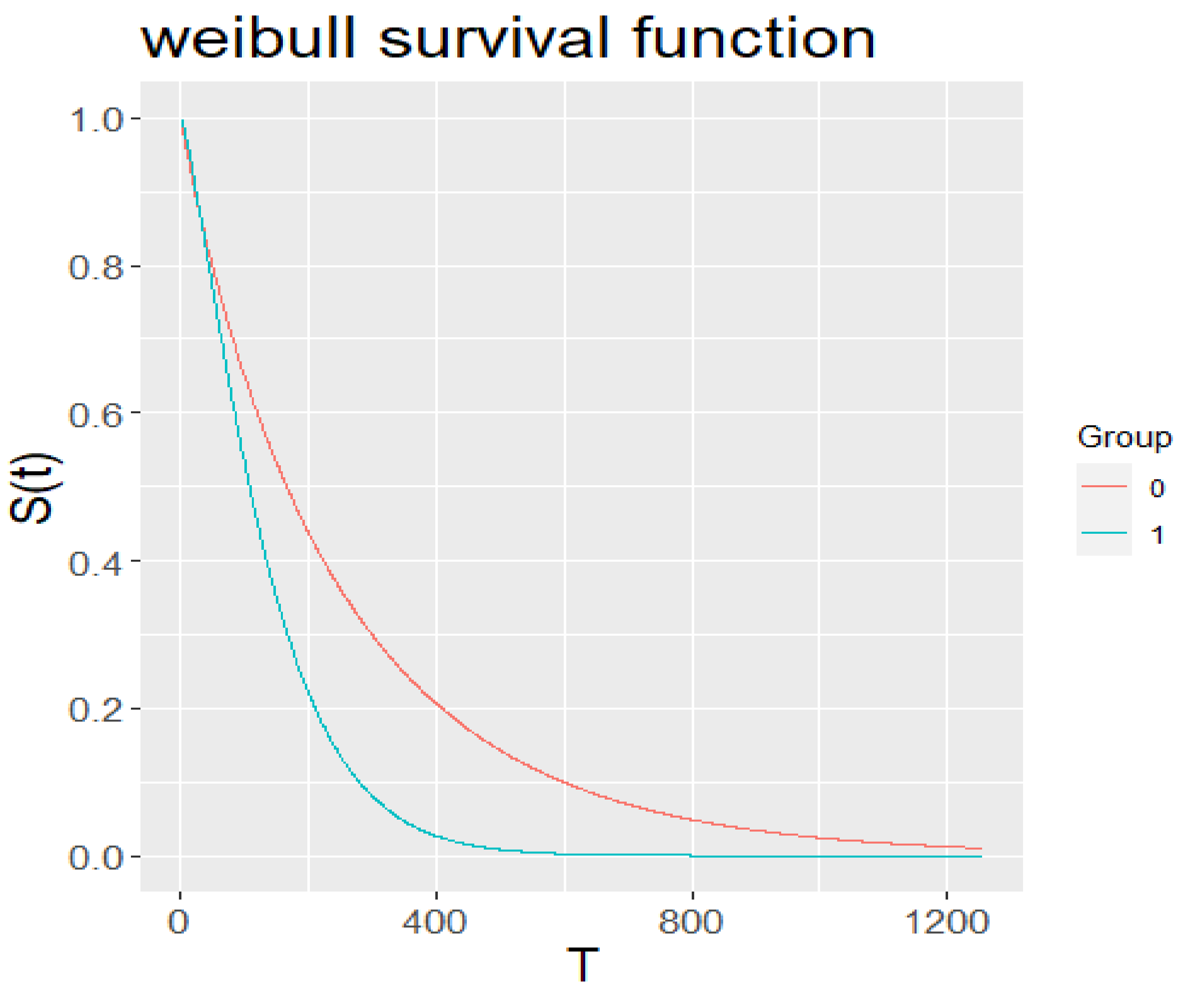
Οι δύο κατανομές (Weibull και Log-Logistic) προσαρμόζονται στα δεδομένα της 2ης ομάδας (Group = 1), με βάση το κριτήριο του AIC.

Το μικρότερο AIC υποδεικνύει καλύτερη προσαρμογή του μοντέλου στα δεδομένα. Συνεπώς, από τα παραπάνω αποτελέσματα, φαίνεται ότι το μοντέλο Log-Logistic Distribution έχει μικρότερη τιμή AIC (388.5781) σε σύγκριση με το μοντέλο Weibull Distribution (391.3207).

Συνεπώς, μπορούμε να συμπεράνουμε ότι το μοντέλο Log-Logistic Distribution προσαρμόζεται καλύτερα στα δεδομένα της 2ης ομάδας σε σχέση με το μοντέλο Weibull Distribution, βάσει του κριτηρίου του AIC.

```
> plot_surv(cc, "weibull",  time = "t", censor = "c", by = "Group")
```

Μεθοδος μεγιστης πιθανοφανειας



iv. **Αν θεωρήσουμε ότι το μοντέλο της Λογαριθμο-λογιστικής κατανομής ταιριάζει στα δεδομένα μας, βρείτε για κάθε ομάδα ασθενών, τις ε.μ.π. των παραμέτρων του μοντέλου αυτού και ακολούθως την ε.μ.π. της διαμέσου χωριστά για τις δύο ομάδες. Με βάση τα αποτελέσματα αυτά, να κριθεί ποιά ομάδα αργεί περισσότερο να αποσυνδεθεί από το οξυγόνο.**

Με αυτήν την παραμέτρηση

$$f(t) = \frac{\gamma\left(\frac{t}{\theta}\right)^{\gamma}}{t\left[1+\left(\frac{t}{\theta}\right)^{\gamma}\right]^{2}},\ t>0,\gamma>0,\theta>0,$$

$$S(t) = \left[1+\left(\frac{t}{\theta}\right)^{\gamma}\right]^{-1}$$

Shape= γ=1/τ
scale=θ=exp(ν)

Ομάδα =0

```
> modllog0<-fit_data(data = ccdat0, dist = "llogis", time = "t", censor =
"c")
> summary(modllog0)
Fitting of the distribution ' llogis ' By maximum likelihood on censored
data
Parameters
        estimate Std. Error
shape   1.341917  0.1729928
scale 138.446997 27.8095493
Loglikelihood:  -261.1803   AIC:  526.3605   BIC:  529.8829
Correlation matrix:
             shape       scale
shape  1.00000000 -0.02933534
scale -0.02933534  1.00000000
```

ΕΜΠ:

Shape= γ=1/τ=1.341917
scale=θ=exp(ν)= 138.446997

$$f(t) = \frac{1.341917\left(\frac{t}{138.446997}\right)^{1.341917}}{t\left[1+\left(\frac{t}{138.446997}\right)^{1.341917}\right]^{2}}$$

$$S(t) = \left[1+\left(\frac{t}{138.446997}\right)^{1.341917}\right]^{-1}$$

Διαμεσος :

$$\boldsymbol{S(L_{50})} = \left[\mathbf{1} + \left(\frac{\boldsymbol{x}}{\mathbf{138.446997}}\right)^{1.341917}\right]^{-\mathbf{1}} = \mathbf{0.5} \iff \boldsymbol{L_{50}} = 138.446997$$

Ομάδα =1

```
> modllog1<-fit_data(data = ccdat1, dist = "llogis", time = "t", censor =
"c")
> summary(modllog1)
Fitting of the distribution ' llogis ' By maximum likelihood on censored
data
Parameters
       estimate Std. Error
shape  1.953817  0.2782326
scale 92.627407 14.1293125
Loglikelihood:  -192.2891   AIC:  388.5781   BIC:  391.6888
Correlation matrix:
            shape       scale
shape  1.00000000 -0.05036393
scale -0.05036393  1.00000000
```

ΕΜΠ:

Shape= γ=1/τ=**1.953817**

scale=θ=exp(ν=**92.627407**

$$f(t)=\frac{1.953817\left(\frac{t}{92.627407}\right)^{1.953817}}{t\left[1+\left(\frac{t}{92.627407}\right)^{1.953817}\right]^{2}}$$

$$S(t)=\left[1+\left(\frac{t}{92.627407}\right)^{1.953817}\right]^{-1}$$

Διαμεσος :

$$S(L_{50})=\left[1+\left(\frac{L_{50}}{92.627407}\right)^{1.953817}\right]^{-1}=0.5\Leftrightarrow L_{50}=92.627407$$