ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

Εργασία 2

Μάθημα: Μοντέλα Αξιοπιστίας και Επιβίωσης

Διδάσκων: Χρυσηίς Καρώνη

Φοιτήτρια: Ελένη Στυλιανού, ge21708

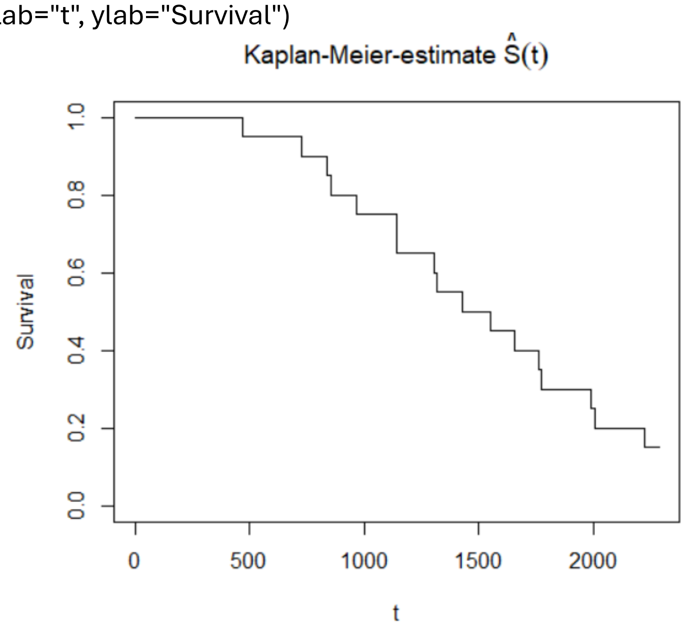
Email: elenistylianou03@live.com

B) Αρχικά, φορτώνουμε στην R τις απαραίτητες βιβλιοθήκες και περνάμε τα δεδομένα επιβίωσης με δεξιά αποκοπή δίνοντας τις εντολές:

```
> library(survival)
> times <- c(468, 725, 838, 853, 965, 1139, 1142, 1304, 1317, 1427,
+         1554, 1658, 1764, 1776, 1990, 2010, 2224, 2244, 2279, 2286)
> censored <- c(rep(0, 17), rep(1, 3))
```

Για την εύρεση των εκτιμήσεων Kaplan-Meier και την δημιουργία γραφικής παραστάσεως αυτών δίνουμε τις εντολές:

```
> surv_obj <- Surv(times, event = 1 - censored)
> km_fit <- survfit(surv_obj ~ 1)
> plot(km_fit, main=expression(paste("Kaplan-Meier-estimate ", hat(S)(t))),
+ conf.int=FALSE,xlab="t", ylab="Survival")
```



Για τις εκτιμήσεις Kaplan–Meier δίνουμε την εντολή:

```
> km_fit$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
```

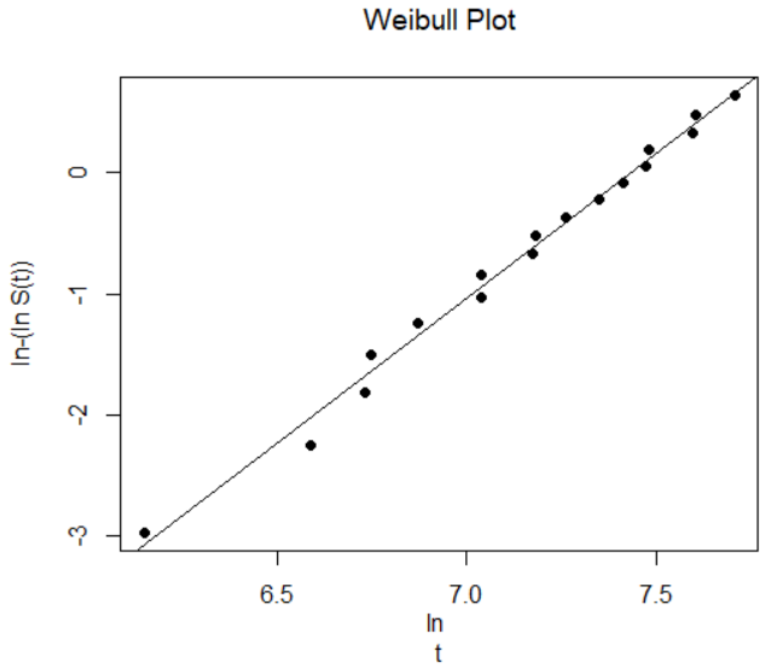
Για τους γραφικούς ελέγχους για τις κατανομές Weibull, Λογαριθμο–κανονική και Λογαριθμο–λογιστική κάνοντας χρήση των εκτιμήσεων Kaplan–Meier $S(t_i)$ δίνουμε τις εντολές:

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

Weibull:

```
>  plot(log(Utime),log(-log(SKM)), main=expression(paste("Weibull Plot")), xlab="ln
+ t", ylab="ln-(ln S(t))", pch=19)
```

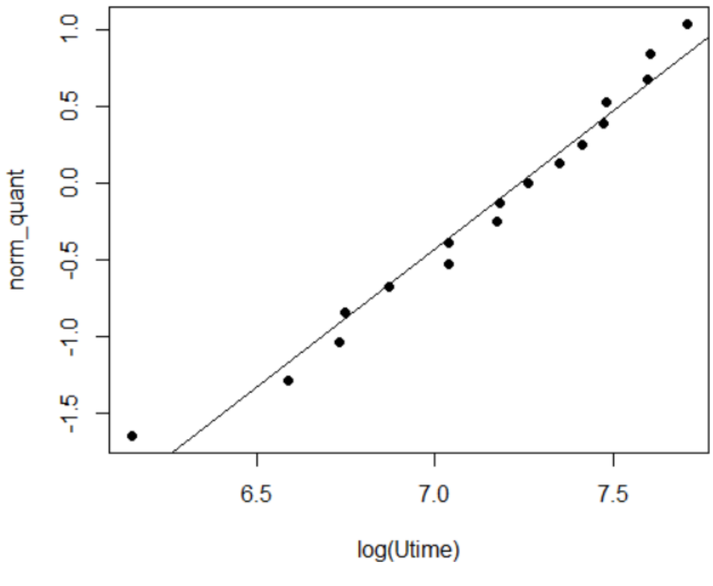
> abline(lm(log(-log(SKM))~log(Utime)))



Από το παραπάνω γράφημα παρατηρούμε ότι υπάρχει σχεδόν τέλεια ευθυγράμμιση στα δεδομένα με την κατανομή, ειδικά για μεγάλους χρόνους γεγονός που υποδηλώνει καλή προσαρμογή στις ουρές. Επομένως, το μοντέλο Weibull ταιριάζει πολύ καλά, ιδίως στις ουρές και στο κύριο σώμα της κατανομής.

Log-Normal:

```
> 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)))
```

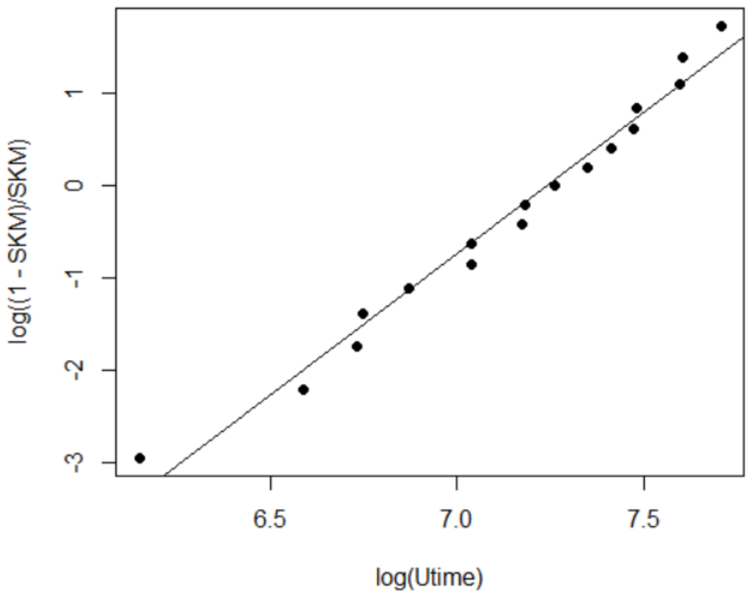


Σε αυτό το γράφημα παρατηρούμε ότι η ευθυγράμμιση είναι πολύ καλή, παρουσιάζεται μια ελαφριά απόκλιση στις ακραίες τιμές. Συνεπώς, έχουμε εξίσου καλή προσαρμογή log-normal, με ελαφρώς μικρότερη ακρίβεια στις ουρές σε σχέση με το Weibull.

Log-Logistic

```
> plot(log(Utime),log((1-SKM)/SKM), pch=19)
```

```
> abline(lm(log((1-SKM)/SKM) ~ log(Utime)))
```



Σε αυτό το γράφημα παρατηρούμε ότι η ευθυγράμμιση των σημείων με την ευθεία είναι πολύ καλή, γεγονός που υποδεικνύει ότι η Log-Logistic κατανομή προσαρμόζεται ικανοποιητικά στα δεδομένα. Υπάρχει μια ελαφριά απόκλιση στις ακραίες τιμές (ουρές), γεγονός που δείχνει ελαφρώς μικρότερη ακρίβεια στις περιοχές αυτές. Συνολικά, όμως, η Log-Logistic κατανομή προσφέρει πολύ καλή προσαρμογή, παρόμοια με αυτή της Weibull, με μικρές διαφορές στην απόδοση στις ουρές.

Για προσαρμογή στις κατανομές Weibull, Λογαριθμο–κανονική και Λογαριθμο–λογιστική δίνουμε τις εξής εντολές:

```
> fit_weibull <- survreg(surv_obj ~ 1, dist = "weibull")
> fit_lognorm <- survreg(surv_obj ~ 1, dist = "lognormal")
> fit_loglogis <- survreg(surv_obj ~ 1, dist = "loglogistic")
```

Για την εύρεση των μοναδικών χρόνων αποτυχίας και τις εκτιμήσεις Kaplan–Meier σε αυτούς δίνουμε τις εντολές:

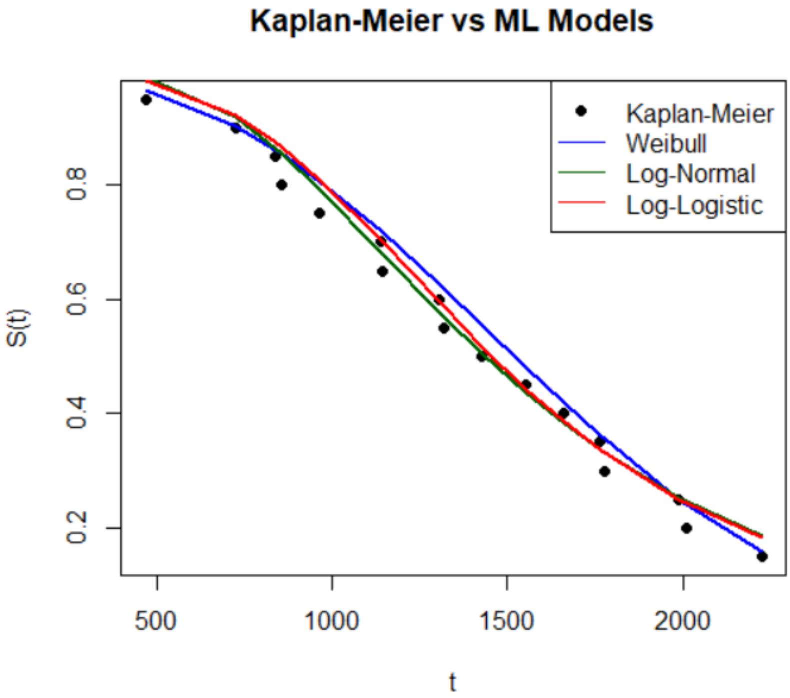
```
> ti <- km_fit$time[km_fit$n.event == 1]
> SKM <- km_fit$surv[km_fit$n.event == 1]
```

Για την εύρεση της συνάρτησης επιβίωσης σε κάθε μοντέλο και την αξιολόγηση της στα σημεία ti δίνουμε τις πιο κάτω εντολές:

Weibull:

```
> mu_w <- fit_weibull$coefficients
> sigma_w <- fit_weibull$scale
> S_weibull <- 1 - pweibull(ti, shape = 1/sigma_w, scale = exp(mu_w))
```

Log-Normal:

```
> mu_ln <- fit_lognorm$coefficients
> sigma_ln <- fit_lognorm$scale
> S_lognorm <- 1 - plnorm(ti, meanlog = mu_ln, sdlog = sigma_ln)
```
Log-Logistic
```
> mu_ll <- fit_loglogis$coefficients
> sigma_ll <- fit_loglogis$scale
> S_loglogis <- 1 / (1 + (ti / exp(mu_ll))^(1 / sigma_ll))
```
Και κάνουμε το γράφημα για τους γραφικούς ελέγχους για τις κατανομές Weibull, Λογαριθμο–κανονική και Λογαριθμο–λογιστική κάνοντας χρήση των εκτιμήσεων Kaplan–Meier $S(t_i)$ δίνοντας τις εντολές:
```
> plot(ti, SKM, pch=19, xlab="t", ylab="S(t)", main="Kaplan-Meier vs ML Models")
> lines(ti, S_weibull, col="blue", lwd=2, type="l")
> lines(ti, S_lognorm, col="darkgreen", lwd=2, type="l")
> lines(ti, S_loglogis, col="red", lwd=2, type="l")
> legend("topright", legend=c("Kaplan-Meier", "Weibull", "Log-Normal", "Log-Logistic"),
+       col=c("black", "blue", "darkgreen", "red"), lty=c(NA,1,1,1), pch=c(19, NA, NA, NA))
```


Από το παραπάνω γράφημα παρατηρούμε ότι όλες οι κατανομές ταιριάζουν σχετικά καλά με την Kaplan–Meier, δηλαδή η προσέγγιση είναι καλή συνολικά. Η Weibull (μπλε γραμμή) φαίνεται να ακολουθεί αρκετά πιστά την Kaplan–Meier σε όλο το εύρος των τιμών, κυρίως για μεσαίους και μεγάλους χρόνους. Έχει πολύ καλή προσαρμογή στις "ουρές". Η Log-Logistic (κόκκινη γραμμή) επίσης παρουσιάζει πολύ καλή προσαρμογή, σχεδόν εφάπτεται με την Kaplan–Meier στα περισσότερα σημεία παρόλο που έχει υπερεκτίμηση σε κάποιες περιοχές. Η Log-Normal (πράσινη) τείνει να υποεκτιμά την επιβίωση για μεσαίους χρόνους και υπερεκτιμά για μικρούς, δηλαδή έχει μεγαλύτερη απόκλιση σε σχέση με τις άλλες δύο κατανομές. Επομένως, παρατηρούμε ότι η Weibull έχει καλύτερη συμπεριφορά στις ουρές και η Log-Logistic έχει καλή ισορροπία γεγονός που τις καθιστά καλά υποψήφια μοντέλα ενώ η Log-Normal έχει την χειρότερη προσαρμογή οπτικά από τις 3 αλλά και αυτή είναι ένα καλό υποψήφιο μοντέλο.

Για τον έλεγχο Anderson–Darling για κάθε μια από αυτές τις κατανομές δίνουμε τις εντολές:

> library(goftest)

> library(fitdistrplus)

> times_full <- times[censored == 0]

Weibull:

> ad.test(times_full, null = function(x) pweibull(x, shape = 1/sigma_w, scale = exp(mu_w)))

```
        Anderson-Darling test of goodness-of-fit
        Null hypothesis: distribution 'function(x) pweibull(x, shape =
        1/sigma_w, scale = exp(mu_w))'
        Parameters assumed to be fixed

data:  times_full
An = 0.91984, p-value = 0.4006
```

Log-Normal:

> ad.test(times_full, null = function(x) plnorm(x, meanlog = mu_ln, sdlog = sigma_ln))

```
        Anderson-Darling test of goodness-of-fit
        Null hypothesis: distribution 'function(x) plnorm(x, meanlog = mu_ln,
        sdlog = sigma_ln)'
        Parameters assumed to be fixed

data:  times_full
An = 0.76821, p-value = 0.5024
```

Log-Logistic:

> ad.test(times_full, null = function(x) {

+ 1 / (1 + (x / exp(mu_ll))^(1 / sigma_ll))

+ })

```
        Anderson-Darling test of goodness-of-fit
        Null hypothesis: distribution 'function(x) {'
        Null hypothesis: distribution ' 1/(1 + (x/exp(mu_ll))^(1/sigma_ll))'
        Null hypothesis: distribution '}'
        Parameters assumed to be fixed

data:  times_full
An = 0.82148, p-value = 0.4639
```

Από τα πιο πάνω αποτελέσματα παρατηρούμε ότι όλες οι κατανομές έχουν καλή προσαρμογή. Σε όλες τις περιπτώσεις, p-value > 0.05, άρα δεν απορρίπτουμε τη μηδενική υπόθεση ότι τα δεδομένα προέρχονται από αυτές τις κατανομές. Όσο μικρότερο το An, τόσο καλύτερη προσαρμογή. Εδώ, η Log-Normal έχει το χαμηλότερο An (0.768), ακολουθεί η Log-Logistic (0.821), και έπειτα η Weibull (0.919). Άρα, με βάση AD, μικρό πλεονέκτημα έχει η Log-Normal, αλλά όλες ταιριάζουν καλά.

Για την εφαρμογή του του κριτηρίου AIC δίνουμε τις εντολές:

> AIC(fit_weibull)

[1] 277.2924

> AIC(fit_lognorm)

[1] 277.1472

> AIC(fit_loglogis)

[1] 277.2844

Από τα πιο πάνω αποτελέσματα παρατηρούμε ότι τελικά η Log-Normal κατανομή έχει το μικρότερο AIC και το μικρότερο AD statistic, άρα με στατιστικά κριτήρια είναι η επικρατέστερη. Η Log-Logistic είναι πολύ κοντά και στα δύο κριτήρια , ελαφρώς χειρότερη αλλά πολύ καλή επιλογή. Η Weibull έχει το μεγαλύτερο AIC και AD, όμως έχει καλύτερη οπτική προσαρμογή στις ουρές, κάτι που μπορεί να είναι σημαντικό αν τα tail probabilities είναι κρίσιμα στο πρόβλημα. Επομένως, για στατιστικά βέλτιστο μοντέλο επιλέγουμε την Log-Normal. Σε περίπτωση μεγάλων χρόνων επιβίωσης ίσως προτιμήσουμε την Weibull. Η Log-Logistic είναι μια καλή προσαρμογή και στα 2.

Για την εκτέλεση του στατιστικού ελέγχου υποθέσεων της Εκθετικής κατανομής (η=1) με εναλλακτική την κατανομή Weibull (η≠1) δίνουμε τις πιο κάτω εντολές:

Προσαρμογή Weibull:

> fit_weibull <- survreg(surv_obj ~ 1, dist = "weibull")

Υπολογισμός η:

> eta_hat <- 1 / fit_weibull$scale

> eta_hat

[1] 2.576594

Για να φτιάξουμε 95% Δ.Ε. για την scale, πρώτα παίρνουμε το standard error:

> se_scale <- summary(fit_weibull)$table["Log(scale)", "Std. Error"]

Παίρνουμε το SE του scale μέσω της κανονικής μετατροπής log:

> scale_log <- log(fit_weibull$scale)

> CI_log <- scale_log + c(-1, 1) * qnorm(0.975) * se_scale

Δημιουργία Δ.Ε. για scale:

> CI_scale <- exp(CI_log)

Υπολογισμός Δ.Ε. για το η:

> CI_eta <- 1 / rev(CI_scale) # αντιστροφή και ανάποδα (γιατί το rev αλλάζει τη σειρά)

> CI_eta

[1] 1.731214 3.834786

Παρατηρούμε ότι το 1 δεν ανήκει στο Δ.Ε. επομένως απορρίπτουμε την εκθετική κατανομή και προτιμάμε την Weibull.

Επίσης, για επιβεβαίωση του αποτελέσματος μέσω του p-value εκτελούμε και τις εξής εντολές:

```
> fit_exp <- survreg(surv_obj ~ 1, dist = "exponential")

> lrt_stat <- 2 * (logLik(fit_weibull) - logLik(fit_exp))

> lrt_stat

'log Lik.' 14.84079 (df=2)

> pchisq(lrt_stat, df = 1, lower.tail = FALSE)

'log Lik.' 0.0001169781 (df=2)
```

Παρατηρούμε ότι το p-value < 0.05 επομένως επιβεβαιώνουμε το συμπέρασμα ότι τα δεδομένα δεν ακολουθούν εκθετική κατανομή αλλά την κατανομή Weibull.

Γ) Αρχικά φορτώνουμε στην R τις απαραίτητες βιβλιοθήκες και τα δεδομένα μας επιλέγοντας το αρχείο «lung-patients.txt» δίνοντας τις εντολές:

```
> library(survival)
> data<- read.table(file.choose(), header=TRUE)
> attach(data)
> data
  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	74	1	56	1
74	75	1	310	0
75	76	1	108	1
76	77	1	51	1

77 78 1 70 1

Για τον υπολογισμό των εκτιμητριών Kaplan–Meier και των γραφικών παραστάσεων αυτών για τις δύο ομάδες δίνουμε τις εντολές:

```
> outp<-survfit(Surv(t,c)~Group, type="kaplan-meier",data=data)
> summary(outp)
```

```
Call: survfit(formula = Surv(t, c) ~ Group, data = data, type = "kaplan-meier")
```

Group=0

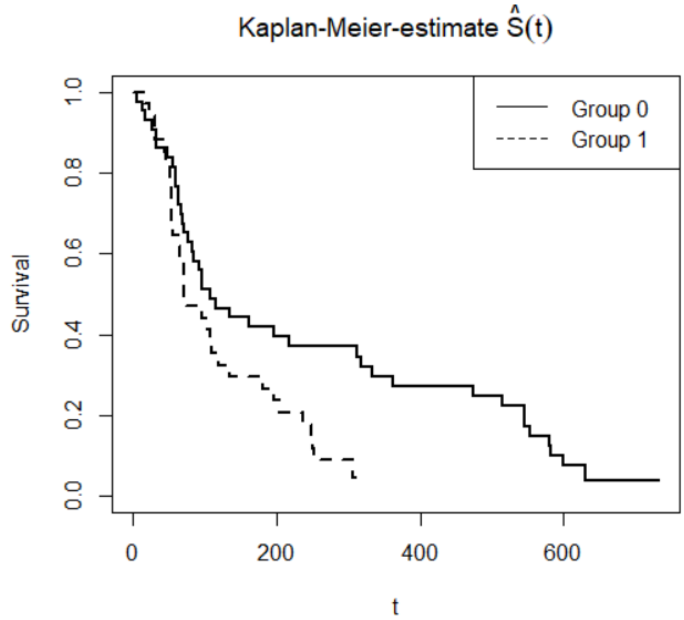
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
5	43	1	0.9767	0.0230	0.93272	1.000
13	42	1	0.9535	0.0321	0.89258	1.000
16	41	1	0.9302	0.0388	0.85712	1.000
26	40	1	0.9070	0.0443	0.82418	0.998
32	39	1	0.8837	0.0489	0.79292	0.985
33	38	1	0.8605	0.0528	0.76289	0.971
48	37	1	0.8372	0.0563	0.73383	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	34	1	0.9706	0.0290	0.91543	1.000
23	33	1	0.9412	0.0404	0.86532	1.000
30	32	1	0.9118	0.0486	0.82124	1.000
31	31	1	0.8824	0.0553	0.78044	0.998
43	30	1	0.8529	0.0607	0.74183	0.981
45	29	1	0.8235	0.0654	0.70486	0.962
51	28	2	0.7647	0.0727	0.63463	0.921
53	26	1	0.7353	0.0757	0.60100	0.900
54	25	2	0.6765	0.0802	0.53616	0.853
56	23	1	0.6471	0.0820	0.50481	0.829
63	22	1	0.6176	0.0833	0.47412	0.805
64	21	1	0.5882	0.0844	0.44403	0.779
70	20	2	0.5294	0.0856	0.38562	0.727
71	18	2	0.4706	0.0856	0.32946	0.672
96	16	1	0.4412	0.0852	0.30222	0.644
103	15	1	0.4118	0.0844	0.27553	0.615
108	14	1	0.3824	0.0833	0.24942	0.586
110	13	1	0.3529	0.0820	0.22390	0.556
118	12	1	0.3235	0.0802	0.19899	0.526
134	11	1	0.2941	0.0781	0.17473	0.495
181	10	1	0.2647	0.0757	0.15117	0.464
197	9	1	0.2353	0.0727	0.12836	0.431
204	8	1	0.2059	0.0693	0.10639	0.398
236	7	1	0.1765	0.0654	0.08537	0.365
249	6	1	0.1471	0.0607	0.06545	0.330
250	5	1	0.1176	0.0553	0.04686	0.295
253	4	1	0.0882	0.0486	0.02995	0.260
306	2	1	0.0441	0.0396	0.00761	0.256

```
> 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)



Για την διεξαγωγή των ελέγχων log–rank και Wilcoxon για τη σύγκριση των δύο ομάδων ως προς τη διάρκεια παροχής οξυγόνου δίνουμε τις πιο κάτω εντολές:

log–rank:

> 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     49.2      1.72       6.2
Group=1 34       32     22.8      3.72       6.2

 Chisq= 6.2  on 1 degrees of freedom, p= 0.01
```

Wilcoxon:

> out2<-survdiff(Surv(t, c) ~ Group, rho=1)

> out2

```
Call:
survdiff(formula = Surv(t, c) ~ Group, rho = 1)

         N Observed Expected (O-E)^2/E (O-E)^2/V
Group=0 43     19.5     23.8     0.792      3.03
Group=1 34     19.2     14.9     1.267      3.03

 Chisq= 3  on 1 degrees of freedom, p= 0.08
```

Από τον έλεγχο log–rank παρατηρούμε ότι p-value = 0.01<0.05, γεγονός που μας δείχνει ότι υπάρχει στατιστικά σημαντική διαφορά μεταξύ των δύο ομάδων. Επίσης, η ομάδα χωρίς ενισχυμένη θεραπεία (Group=1) είχε λιγότερα αναμενόμενα συμβάντα (22.8) αλλά 32 παρατηρηθέντα, άρα χειρότερη επιβίωση. Επομένως, η ομάδα με ενισχυμένη θεραπεία (Group=0) έχει σημαντικά μεγαλύτερη διάρκεια παροχής οξυγόνου.

Από τον έλεγχο Wilcoxon παρατηρούμε ότι p-value = 0.08>0.05, γεγονός που μας δείχνει ότι δχεν υπάρχει στατιστικά σημαντική διαφορά. Αυτό μπορεί να οφείλεται στο γεγονός ότι αυτός ο έλεγχος δίνει περισσότερο βάρος στα πρώιμα γεγονότα. Η οριακή μη σημαντικότητα του αποτελέσματος μα δείχνει ότι η διαφορά ίσως να μην είναι τόσο έντονη στην αρχή, αλλά συνολικά υπάρχει σαφής ένδειξη υπεροχής της ενισχυμένης θεραπείας.

Για τους γραφικούς ελέγχους για τις κατανομές Weibull και Λογαριθμο–λογιστική κάνοντας χρήση των εκτιμήσεων Kaplan–Meier $S(t_i)$ για τις δύο ομάδες χωριστά δίνουμε τις εντολές:

Weibull:

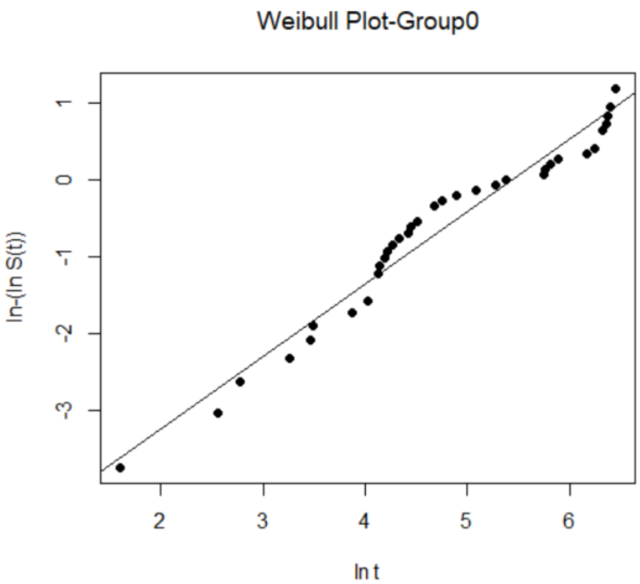
Για το Group 0:

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

```
Call: survfit(formula = group0 ~ 1, data = data, 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
```

```
> Utime0<-outp0$time[outp0$n.event==1]
> SKM0<-outp0$surv[outp0$n.event==1]
> plot(log(Utime0),log(-log(SKM0)) , main=expression(paste("Weibull Plot-Group0")), xlab="ln t", ylab="ln-(ln S(t))", pch=19)
> abline(lm(log(-log(SKM0))~log(Utime0)))
```



Από την πιο πάνω γραφική παράσταση παρατηρούμε ότι υπάρχει απόκλιση από την ευθεία, γεγονός που μας δείχνει πως το μοντέλο δε εφαρμόζεται τόσο καλά σε αυτήν την ομάδα, επομένως δεν περιγράφει επαρκώς τα δεδομένα αυτής της ομάδας.

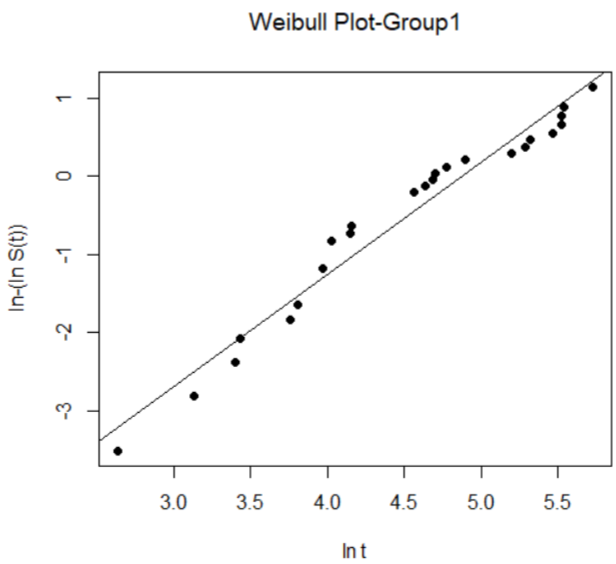
Για το Group 1:

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

```
Call: survfit(formula = group1 ~ 1, data = data, type = "kaplan-meier")

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   14     34       1   0.9706  0.0290      0.91543        1.000
   23     33       1   0.9412  0.0404      0.86532        1.000
   30     32       1   0.9118  0.0486      0.82124        1.000
   31     31       1   0.8824  0.0553      0.78044        0.998
   43     30       1   0.8529  0.0607      0.74183        0.981
   45     29       1   0.8235  0.0654      0.70486        0.962
   51     28       2   0.7647  0.0727      0.63463        0.921
   53     26       1   0.7353  0.0757      0.60100        0.900
   54     25       2   0.6765  0.0802      0.53616        0.853
   56     23       1   0.6471  0.0820      0.50481        0.829
   63     22       1   0.6176  0.0833      0.47412        0.805
   64     21       1   0.5882  0.0844      0.44403        0.779
   70     20       2   0.5294  0.0856      0.38562        0.727
   71     18       2   0.4706  0.0856      0.32946        0.672
   96     16       1   0.4412  0.0852      0.30222        0.644
  103     15       1   0.4118  0.0844      0.27553        0.615
  108     14       1   0.3824  0.0833      0.24942        0.586
  110     13       1   0.3529  0.0820      0.22390        0.556
  118     12       1   0.3235  0.0802      0.19899        0.526
  134     11       1   0.2941  0.0781      0.17473        0.495
  181     10       1   0.2647  0.0757      0.15117        0.464
  197      9       1   0.2353  0.0727      0.12836        0.431
  204      8       1   0.2059  0.0693      0.10639        0.398
  236      7       1   0.1765  0.0654      0.08537        0.365
  249      6       1   0.1471  0.0607      0.06545        0.330
  250      5       1   0.1176  0.0553      0.04686        0.295
  253      4       1   0.0882  0.0486      0.02995        0.260
  306      2       1   0.0441  0.0396      0.00761        0.256
```

```
> Utime1<-outp1$time[outp1$n.event==1]
> SKM1<-outp1$surv[outp1$n.event==1]
> plot(log(Utime1),log(-log(SKM1)) , main=expression(paste("Weibull Plot-Group1")),
+ xlab="ln t", ylab="ln-(ln S(t))", pch=19)
> abline(lm(log(-log(SKM1))~log(Utime1)))
```



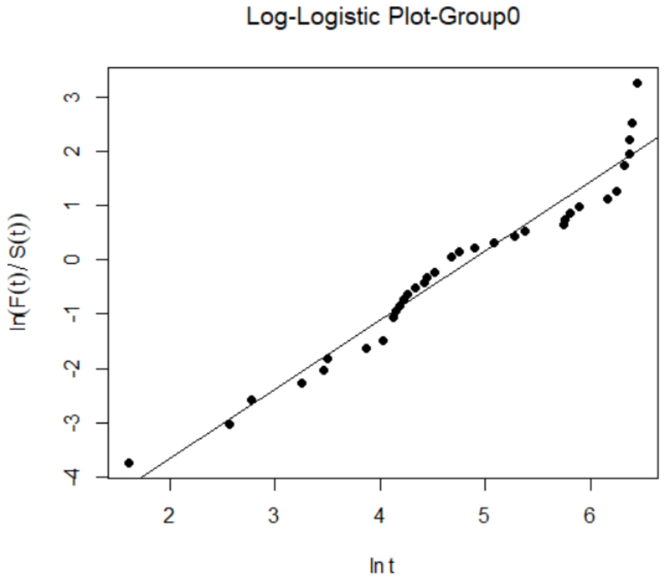
Στο πιο πάνω γράφημα παρατηρούμε ότι τα σημεία είναι πιο ευθυγραμμισμένα με την ευθεία σε σχέση με της άλλης ομάδας. Παρατηρούμε καλύτερη προσαρμογή και ως εκ τούτου είναι πιο πιθανό τα δεδομένα του Group 1 να ακολουθούν κατανομή Weibull.

Log-Logistic:

Για το Group 0:

> plot(log(Utime0),log((1-SKM0)/SKM0), main=expression(paste("Log-Logistic Plot-Group0")), xlab="ln t", ylab=expression(ln (F(t)/S(t))), pch=19)

> abline(lm(log((1-SKM0)/SKM0)~log(Utime0)))

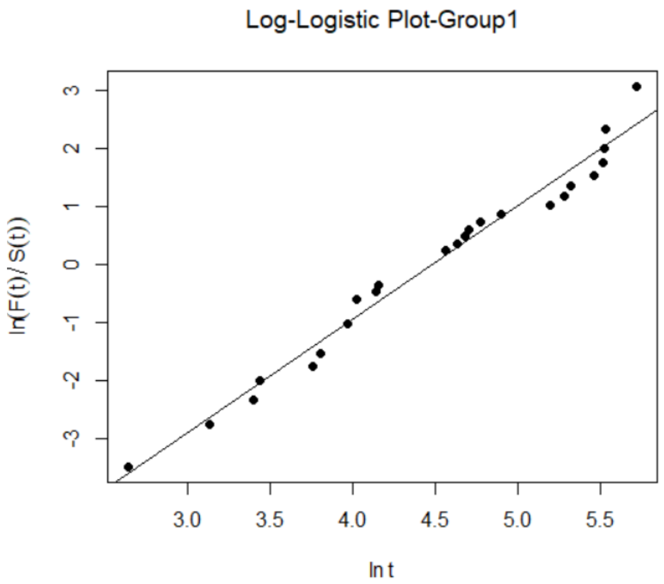


Στο παραπάνω γράφημα παρατηρούμε ότι τα δεδομένα ευθυγραμμίζονται αρκετά καλά με την ευθεία, ωστόσο υπάρχει απόκλιση στα άκρα. Παρόλα αυτά, σε σύγκριση με το γράφημα της κατανομής Weibull η προσαρμογή φαίνεται καλύτερη.

Για το Group 1:

> plot(log(Utime1),log((1-SKM1)/SKM1), main=expression(paste("Log-Logistic Plot-Group1")), xlab="ln t", ylab=expression(ln (F(t)/S(t))), pch=19)

> abline(lm(log((1-SKM1)/SKM1)~log(Utime1)))



Στο πιο πάνω γράφημα παρατηρούμε ότι το γράφημα είναι πανομοιότυπο με αυτό της κατανομής Weibull, γεγονός που μας δείχνει ότι τα δεδομένα του Group 1 μπορούν να προσαρμοστούν και με λογαριθμο-λογιστική κατανομή.

Για επιπρόσθετο έλεγχο κάνουμε χρήση του κριτηρίου AIC δίνοντας τις πιο κάτω εντολές:

```
> fit_weibull_g0 <- flexsurvreg(Surv(t, c) ~ 1, data = data[Group == "0", ], dist = "weibull")
> fit_weibull_g1 <- flexsurvreg(Surv(t, c) ~ 1, data = data[Group == "1", ], dist = "weibull")
> fit_loglogis_g0 <- flexsurvreg(Surv(t, c) ~ 1, data = data[Group == "0", ], dist = "llogis")
> fit_loglogis_g1 <- flexsurvreg(Surv(t, c) ~ 1, data = data[Group == "1", ], dist = "llogis")
> AIC(fit_weibull_g0)
[1] 525.6616
> AIC(fit_weibull_g1)
[1] 374.2082
> AIC(fit_loglogis_g0)
[1] 526.3605
> AIC(fit_loglogis_g1)
[1] 371.5298
```

Με αυτό το κριτήριο παρατηρούμε ότι για το Group 0 έχει μικρότερο AIC η κατανομή Weibull επομένως έχει καλύτερη προσαρμογή στα δεδομένα άλλα τα δύο AIC έχουν αρκετά μικρή διαφορά. Για το το Group 1 έχει μικρότερο AIC η κατανομή Log-Logistic επομένως έχει και καλύτερη προσαρμογή στα δεδομένα της ομάδας αλλά και πάλι είναι πολύ μικρή η διαφορά.

Επομένως, παρατηρούμε ότι και οι δύο ομάδες μπορούν να προσαρμοστούν με μοντέλο λογαριθμο-λογιστικής κατανομής.

Θεωρώντας τώρα το μοντέλο λογαριθμο-λογιστικής κατανομής θα βρούμε τις ε.μ.π. των παραμέτρων του μοντέλου αυτού και την ε.μ.π. της διαμέσου για κάθε ομάδα ασθενών χρησιμοποιώντας τις πιο κάτω εντολές:

```
> fit_loglogis_g0
Call:
flexsurvreg(formula = Surv(t, c) ~ 1, data = data[Group == "0",
    ], dist = "llogis")

Estimates:
       est      L95%     U95%     se
shape    1.342    1.042    1.728    0.173
scale  138.481   93.412  205.296   27.818

N = 43,  Events: 40,  Censored: 3
Total time at risk: 10031
Log-likelihood = -261.1803, df = 2
AIC = 526.3605
```

```
> fit_loglogis_g1
Call:
flexsurvreg(formula = Surv(t, c) ~ 1, data = data[Group == "1",
    ], dist = "llogis")

Estimates:
       est      L95%     U95%     se
shape    2.033    1.530    2.700    0.294
scale   88.751   66.337  118.738   13.181

N = 34,  Events: 32,  Censored: 2
Total time at risk: 4043
Log-likelihood = -183.7649, df = 2
AIC = 371.5298
```

Παρατηρούμε ότι για το Group 0 οι εκτιμήσεις των παραμέτρων shape και scale είναι 1.342 και 138.481 αντίστοιχα. Επιπλέον, για το Group 1 είναι 2.033 και 88.751 αντίστοιχα. Η διάμεσος σε αυτήν την κατανομή ισούται με την παράμετρο scale. Επομένως, οι εκτιμήσεις της διάμεσου για κάθε ομάδα είναι:

- Group 0: 138.48 μονάδες χρόνου
- Group 1: 88.75 μονάδες χρόνου

Συνεπώς, το Group 0 εμφανίζει μεγαλύτερη εκτιμώμενη διάμεσο χρόνου αποσύνδεσης από το οξυγόνο σε σχέση με το Group 1. Αυτό υποδηλώνει ότι οι ασθενείς της ομάδας Group 0 παραμένουν περισσότερο χρόνο συνδεδεμένοι με το οξυγόνο, άρα η ομάδα Group 0 αργεί περισσότερο να αποσυνδεθεί από το οξυγόνο.

>summary(fit_loglogis_g0)

	time	est	lcl	ucl
1	5	0.98854064	0.96537169	0.9973400
2	13	0.95988475	0.90914874	0.9868804
3	16	0.94766767	0.88901887	0.9816220
4	26	0.90420126	0.82651884	0.9591809
5	32	0.87719539	0.79001398	0.9432756
6	33	0.87267690	0.78391460	0.9403022
7	48	0.80564728	0.69927128	0.8991268
8	56	0.77119831	0.65933659	0.8761353
9	59	0.75860613	0.64424540	0.8669435
10	62	0.74620814	0.63041000	0.8579954
11	63	0.74211997	0.62608611	0.8549569
12	66	0.72999223	0.61288266	0.8452072
13	68	0.72202294	0.60428575	0.8389157
14	71	0.71024579	0.59065186	0.8292090
15	76	0.69109540	0.56646446	0.8126638
16	83	0.66529900	0.53594223	0.7886850
17	85	0.65814606	0.52808574	0.7818318
18	91	0.63726265	0.50331834	0.7644256
19	95	0.62381443	0.48767699	0.7524623
20	107	0.58567836	0.44546834	0.7178697
21	116	0.55915676	0.41545529	0.6918048
22	134	0.51103496	0.36875797	0.6480102
23	161	0.44961903	0.31079310	0.5848893
24	193	0.39042909	0.25730513	0.5234266
25	196	0.38551375	0.25281196	0.5183733
26	217	0.35370118	0.22267692	0.4817512
27	313	0.25078711	0.14173921	0.3671174
28	317	0.24759858	0.13911645	0.3625384
29	333	0.23549375	0.12978330	0.3484316
30	362	0.21591861	0.11503678	0.3242328
31	475	0.16054035	0.07735041	0.2566527
32	514	0.14677606	0.06922453	0.2394465
33	546	0.13691304	0.06354899	0.2258253
34	553	0.13490527	0.06240856	0.2229862
35	580	0.12761166	0.05823870	0.2136282
36	583	0.12684283	0.05775724	0.2127803
37	600	0.12263170	0.05517334	0.2074340
38	619	0.11820061	0.05260397	0.2014025
39	631	0.11554109	0.05110759	0.1977536
40	733	0.09652504	0.04000961	0.1741875

> summary(fit_loglogis_g1)

	time	est	lcl	ucl
1	14	0.97710715	0.93729095	0.9941393
2	23	0.93961383	0.87135640	0.9787761
3	30	0.90066522	0.81528030	0.9584413
4	31	0.89454104	0.80518399	0.9550683
5	43	0.81349665	0.69072665	0.9040202
6	45	0.79906893	0.67215832	0.8935142
7	51	0.75511427	0.61840073	0.8588659
8	53	0.74036960	0.60217629	0.8469303
9	54	0.73299992	0.59473032	0.8407685
10	56	0.71828598	0.57692823	0.8300969
11	63	0.66742649	0.52464246	0.7892917
12	64	0.66028355	0.51801881	0.7827943
13	70	0.61831715	0.47412628	0.7438954
14	71	0.61148999	0.46686326	0.7381587
15	96	0.46018585	0.31913794	0.5935877
16	103	0.42490950	0.28165707	0.5595070
17	108	0.40155227	0.25772483	0.5353255
18	110	0.39262340	0.24983666	0.5252551
19	118	0.35916364	0.22190020	0.4947176
20	134	0.30207149	0.17701256	0.4375280
21	181	0.19022144	0.09535429	0.3105230
22	197	0.16510178	0.07940383	0.2793025
23	204	0.15554993	0.07333060	0.2665920
24	236	0.12047983	0.05252119	0.2206926
25	249	0.10939996	0.04574955	0.2064758
26	250	0.10860873	0.04527763	0.2054392
27	253	0.10628357	0.04390039	0.2023769
28	274	0.09184144	0.03567664	0.1816993
29	306	0.07475300	0.02683776	0.1539962
30	310	0.07294756	0.02598527	0.1509708