

Túlélés elemzés

2017. október 27.

Néhány példa

Egy adott betegség diagnózisától kezdve mennyi ideje van hátra a páciensnek? Tipikusan mennyi ideig élhet túl?

Bizonyos ráktípus esetén mennyi idő telik el a relapszusig?

Kimutatható-e, hogy egy adott kezelés szignifikánsan növeli a túlélési időt egy másik kezeléssel szemben?

Hogyan befolyásolja a gyártási folyamat a termék meghibásodásáig eltelt időt? Melyek a gyártási folyamat jellemzői közül a legnagyobb jelentőségűek a meghibásodásig eltelt idő szempontjából?

Túlélés elemzés

Közös jellemzők:

- A fő kimeneti változó az eseményig eltelt idő mértéke
- Nem minden minta esetén történik esemény =>
cenzorált minták
 - *a vizsgálat ideje alatt nem történik esemény az adott minta esetén (a vizsgálat időtartama túlságosan rövid)*
 - *az adott mintát „elveszítjük”, nem követjük tovább*
 - *az adott mintával történik egy másik esemény, ami miatt kiesik a vizsgálatból*
- Az eseményig eltelt idő tipikusan nem normális eloszlású

Túlélés és hazard

Túlélés valószínűsége, túlélési függvény (Survival probability, survival function):

$S(t)$: Annak valószínűsége, hogy egy adott személy a kezdőidőponttól kezdve a t időpontig túlél.

Hazard (hazard function):

$h(t)$: Annak valószínűsége, hogy egy személy a t időpontban elszenved egy eseményt, ha addig nem szenvedett el.

Túlélés valószínűségének becslése: Kaplan-Meier módszer

k számú esemény:

$$t_1 < t_2 < t_3 < t_4 < t_5 < \dots < t_k$$

$$S(t_j) = S(t_{j-1}) \left(1 - \frac{d_j}{n_j}\right), \text{ ahol}$$

n_j = az eseményt nem elszenvedő minták száma a t_j időpont előtt

d_j = az események száma a t_j időpontban

$$t_0 = 0 \text{ és } S(0) = 1$$

Kaplan-Meier túlélési görbe: $S(t_j)$ alakulása az időben

Túlélés elemzés R-ben

Példa

```
library(survival)
data("colon")
View( colon )
summary( colon )
```

id:	id
study:	1 for all patients
rx:	Treatment - Obs(ervation), Lev(amisole), Lev(amisole)+5-FU
sex:	1=male
age:	in years
obstruc t:	obstruction of colon by tumour
perfor:	perforation of colon
adhere:	adherence to nearby organs
nodes:	number of lymph nodes with detectable cancer
time:	days until event or censoring
status:	censoring status
differ:	differentiation of tumour (1=well, 2=moderate, 3=poor)
extent:	Extent of local spread (1=submucosa, 2=muscle, 3=serosa, 4=contiguous structures)
surg:	time from surgery to registration (0=short, 1=long)
node4:	more than 4 positive lymph nodes
etype:	event type: 1=recurrence,2=death

```
colon.death <- subset( colon, etype == 2 )
```

```
fit <- survfit( formula = Surv(time, status) ~ 1,  
               data = colon.death )
```

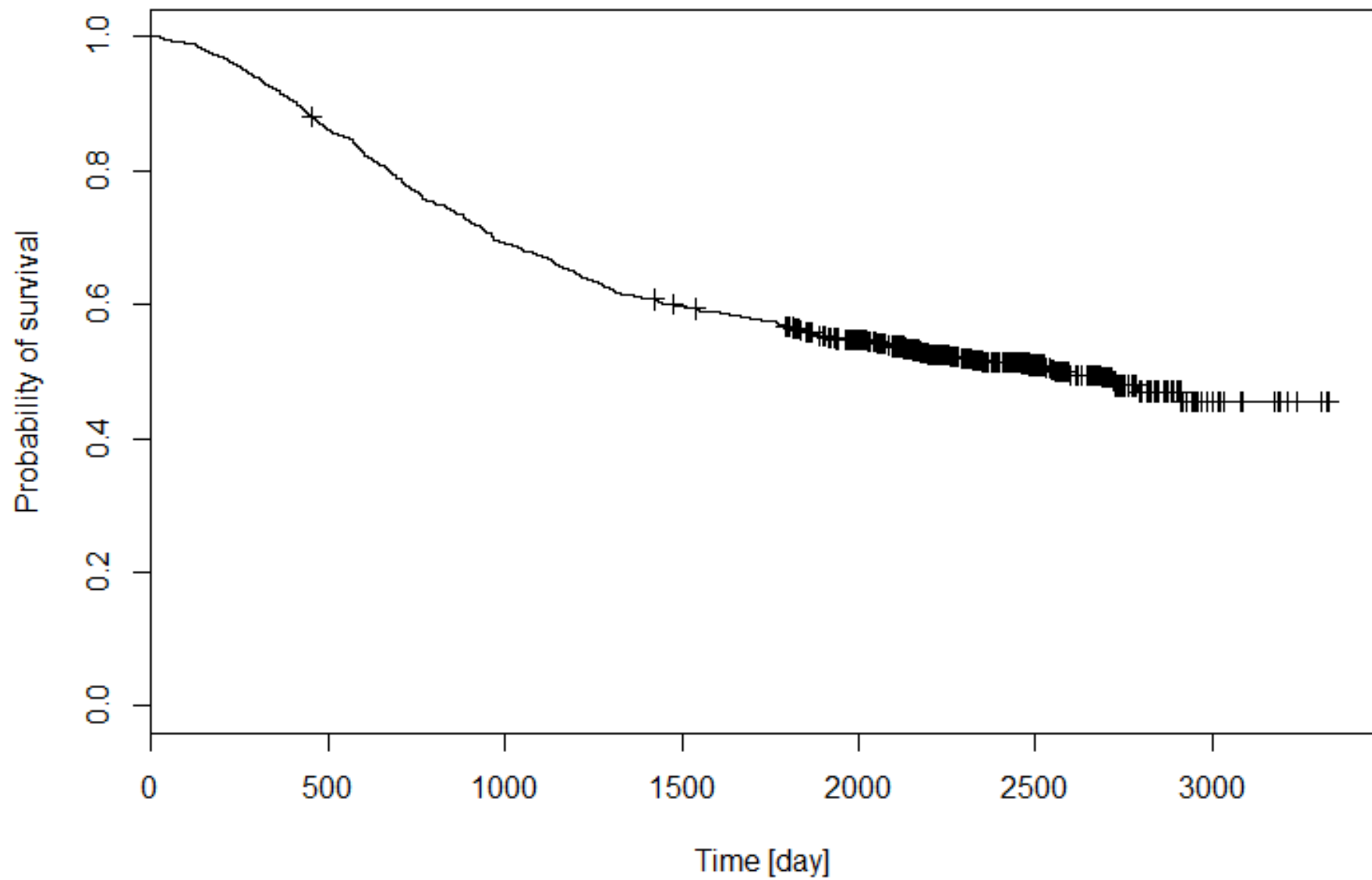
fit

```
Call: survfit(formula = Surv(time, status) ~ 1, data = colon.death)
```

n	events	median	0.95LCL	0.95UCL
929	452	2552	2171	NA

```
plot( fit )
```

```
plot( fit,  
      mark.time = T,  
      conf.int = F,  
      xlab = "Time [day]",  
      ylab = "Probability of survival" )
```



summary(fit)

```
call: survfit(formula = Surv(time, status) ~ 1, data = colon.death)
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
23	929	1	0.999	0.00108	0.997	1.000
24	928	1	0.998	0.00152	0.995	1.000
34	927	1	0.997	0.00186	0.993	1.000
45	926	1	0.996	0.00215	0.991	1.000
52	925	1	0.995	0.00240	0.990	0.999
56	924	1	0.994	0.00263	0.988	0.999
79	923	1	0.992	0.00284	0.987	0.998
93	922	1	0.991	0.00303	0.985	0.997
113	921	1	0.990	0.00321	0.984	0.997
122	920	1	0.989	0.00339	0.983	0.996
125	919	1	0.988	0.00355	0.981	0.995
127	918	1	0.987	0.00370	0.980	0.994
129	917	1	0.986	0.00385	0.978	0.994
133	916	1	0.985	0.00400	0.977	0.993
138	915	1	0.984	0.00414	0.976	0.992
141	914	1	0.983	0.00427	0.974	0.991
144	913	1	0.982	0.00440	0.973	0.990
145	912	1	0.981	0.00452	0.972	0.990
150	911	1	0.980	0.00464	0.970	0.989
164	910	1	0.978	0.00476	0.969	0.988
165	909	1	0.977	0.00488	0.968	0.987
166	908	1	0.976	0.00499	0.967	0.986
171	907	2	0.974	0.00520	0.964	0.984
186	905	1	0.973	0.00531	0.963	0.984
187	904	1	0.972	0.00541	0.961	0.983
191	903	1	0.971	0.00551	0.960	0.982
201	902	1	0.970	0.00561	0.959	0.981
202	901	1	0.969	0.00571	0.958	0.980

Túlélési görbék összehasonlítása

Nemparametrikus teszt: logrank (Peto et al, 1977)

Kettő vagy több csoport összehasonlítása (#csoport = G)

- különböző kezelési ágak
- prognosztikai ágak (pl. súlyosság besorolása)
- ...

Teszt statisztika:

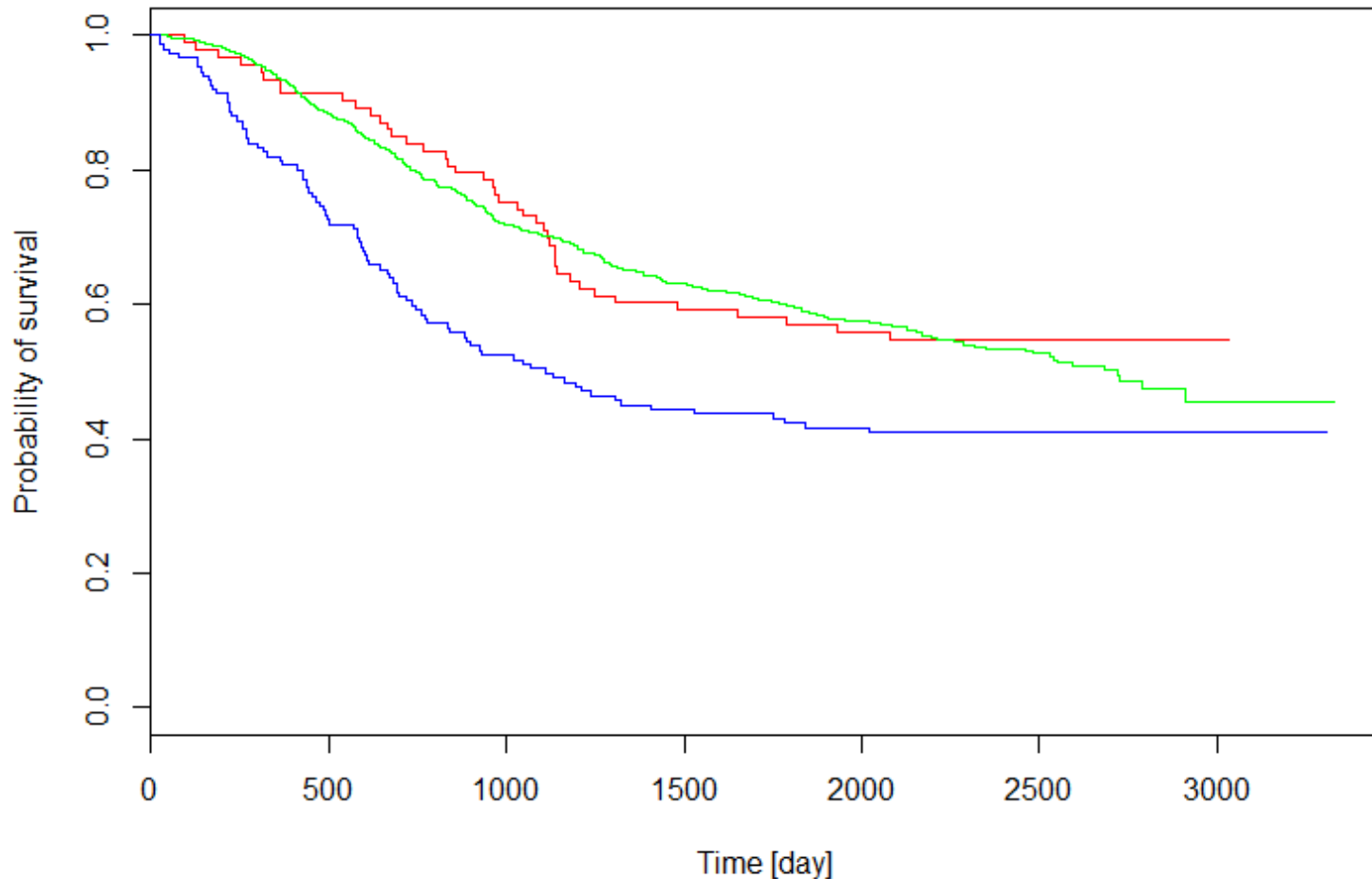
$$X^2 = \sum_{i=1}^G \frac{(O_i - E_i)^2}{E_i}$$

χ^2 eloszlás, (G-1) szabadsági fok

logrank teszt R-ben

```
> survdiff( surv(time, status) ~ differ, data = colon.death )  
Call:  
survdiff(formula = surv(time, status) ~ differ, data = colon.death)  
  
n=906, 23 observations deleted due to missingness.  
  
      N Observed Expected (O-E)^2/E (O-E)^2/V  
differ=1  93      42    47.5      0.643     0.721  
differ=2 663     311   334.9      1.708     7.108  
differ=3 150      88    58.6     14.808    17.105  
  
Chisq= 17.2  on 2 degrees of freedom, p= 0.000185
```

```
plot( survfit( formula = Surv(time, status) ~ differ,  
              data = colon.death ),  
      col = c("red","green","blue"),  
      xlab = "Time [day]",  
      ylab = "Probability of survival" )
```



Feladat

Mely további változók befolyásolják szignifikánsan a túlélési időt?

Ábrázoljuk a Kaplan-Meier görbét!

Többváltozós elemzés

Milyen esetekben van értelme?

1. Adott prediktor változó hatását akarjuk elemezni, de más prognosztikus faktorok is léteznek.
2. Több prediktor változót vizsgálunk, hogy mennyire képesek megjósolni a túlélést.
3. Több prediktor változót vizsgálunk, hogy összefüggésben vannak-e a túléléssel.
4. ...

A prediktorok hatását számszerűsíteni szeretnénk =>
regressziós elemzés

Cox-regresszió

Proportional Hazard modell

$$h(t) = h(t_0) * \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)$$

$h(t)$: hazard függvény

$h(t_0)$: baseline hazard függvény

X_1, \dots, X_n : kovariánsok (változók)

β_1, \dots, β_n : koefficiensek

Kezelés hatása

```
> table(colon.death$rx)
```

obs	Lev	Lev+5FU
315	310	304

```
> fit <- coxph( formula = surv(time, status) ~ rx, data = colon.death )  
> summary( fit )
```

Call:

```
coxph(formula = surv(time, status) ~ rx, data = colon.death)
```

n= 929, number of events= 452

	coef	exp(coef)	se(coef)	z	Pr(> z)
rxLev	-0.02664	0.97371	0.11030	-0.241	0.80917
rxLev+5FU	-0.37171	0.68955	0.11875	-3.130	0.00175 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
rxLev	0.9737	1.027	0.7844	1.2087
rxLev+5FU	0.6896	1.450	0.5464	0.8703

Concordance= 0.536 (se = 0.013)

Rsquare= 0.013 (max possible= 0.998)

Likelihood ratio test= 12.15 on 2 df, p=0.002302

Wald test = 11.56 on 2 df, p=0.003092

Score (logrank) test = 11.68 on 2 df, p=0.002906

Többváltozós modell

```
> fit <- coxph( formula = Surv(time, status) ~ rx + sex + age + perfor + adhere, data = colon.death )  
> summary( fit )
```

Call:

```
coxph(formula = Surv(time, status) ~ rx + sex + age + perfor +  
      adhere, data = colon.death)
```

n= 929, number of events= 452

	coef	exp(coef)	se(coef)	z	Pr(> z)	
rxLev	-0.035927	0.964711	0.110404	-0.325	0.74487	
rxLev+5FU	-0.368299	0.691910	0.118892	-3.098	0.00195	**
sex	0.003380	1.003385	0.094326	0.036	0.97142	
age	0.001907	1.001909	0.004036	0.472	0.63659	
perfor	0.065523	1.067717	0.266077	0.246	0.80548	
adhere	0.290225	1.336728	0.125860	2.306	0.02111	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
rxLev	0.9647	1.0366	0.7770	1.1978
rxLev+5FU	0.6919	1.4453	0.5481	0.8735
sex	1.0034	0.9966	0.8340	1.2071
age	1.0019	0.9981	0.9940	1.0099
perfor	1.0677	0.9366	0.6338	1.7986
adhere	1.3367	0.7481	1.0445	1.7107

Concordance= 0.55 (se = 0.014)

Rsquare= 0.019 (max possible= 0.998)

Likelihood ratio test= 17.83 on 6 df, p=0.006672

Wald test = 17.65 on 6 df, p=0.007169

Score (logrank) test = 17.83 on 6 df, p=0.006661

Tumor differenciációja

```
> table( colon.death$differ)
```

```
  1    2    3  
93 663 150
```

(1=well,
2=moderate,
3=poor)

```
> fit <- coxph( formula = Surv(time, status) ~ differ, data = colon.death )  
> summary( fit )
```

Call:

```
coxph(formula = Surv(time, status) ~ differ, data = colon.death)
```

n= 906, number of events= 441

(23 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z)	
differ	0.32788	1.38803	0.09618	3.409	0.000651	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
differ	1.388	0.7204	1.15	1.676

Concordance= 0.544 (se = 0.011)

Rsquare= 0.013 (max possible= 0.998)

Likelihood ratio test= 11.51 on 1 df, p=0.0006916

wald test = 11.62 on 1 df, p=0.0006515

score (logrank) test = 11.57 on 1 df, p=0.0006689

Kategorikus változóként

```
> fit <- coxph( formula = surv(time, status) ~ factor(differ), data = colon.death )  
> summary( fit )
```

Call:

```
coxph(formula = surv(time, status) ~ factor(differ), data = colon.death)
```

n= 906, number of events= 441

(23 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(> z)
factor(differ)2	0.04963	1.05088	0.16441	0.302	0.76275
factor(differ)3	0.53196	1.70226	0.18764	2.835	0.00458 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
factor(differ)2	1.051	0.9516	0.7614	1.450
factor(differ)3	1.702	0.5875	1.1784	2.459

Concordance= 0.544 (se = 0.011)

Rsquare= 0.017 (max possible= 0.998)

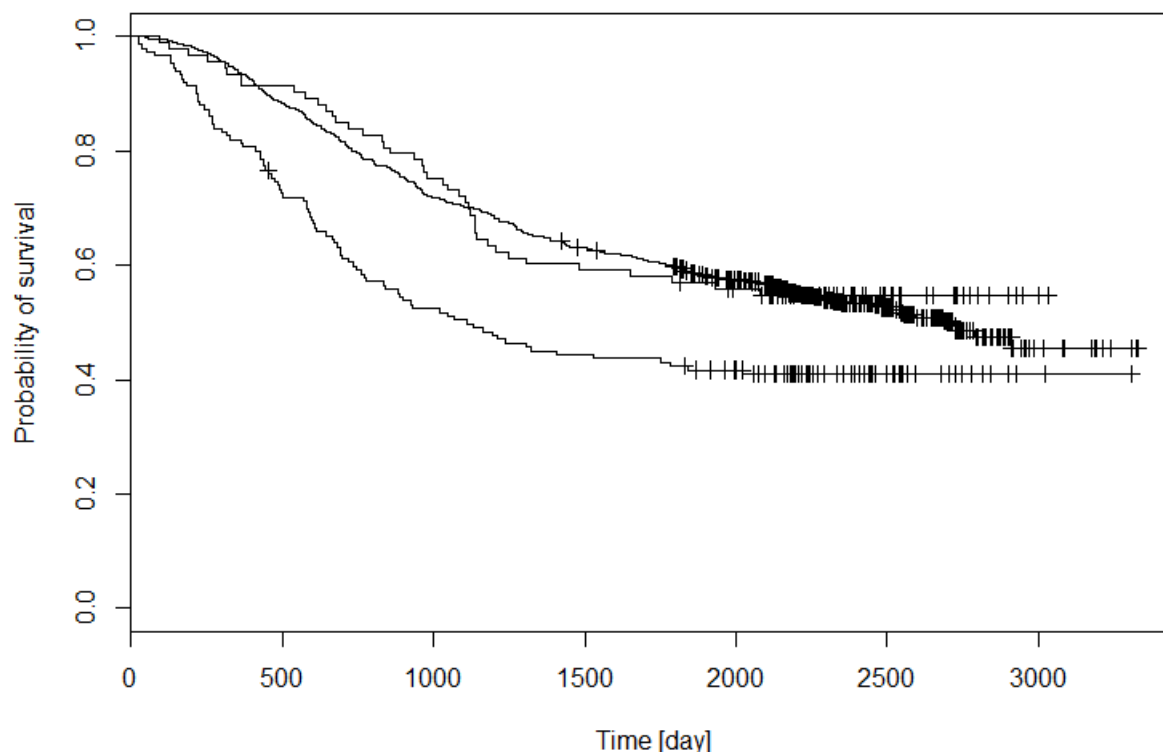
Likelihood ratio test= 15.25 on 2 df, p=0.0004872

Wald test = 16.85 on 2 df, p=0.0002195

Score (logrank) test = 17.19 on 2 df, p=0.0001855

Arányos kockázat fennállásának tesztelése

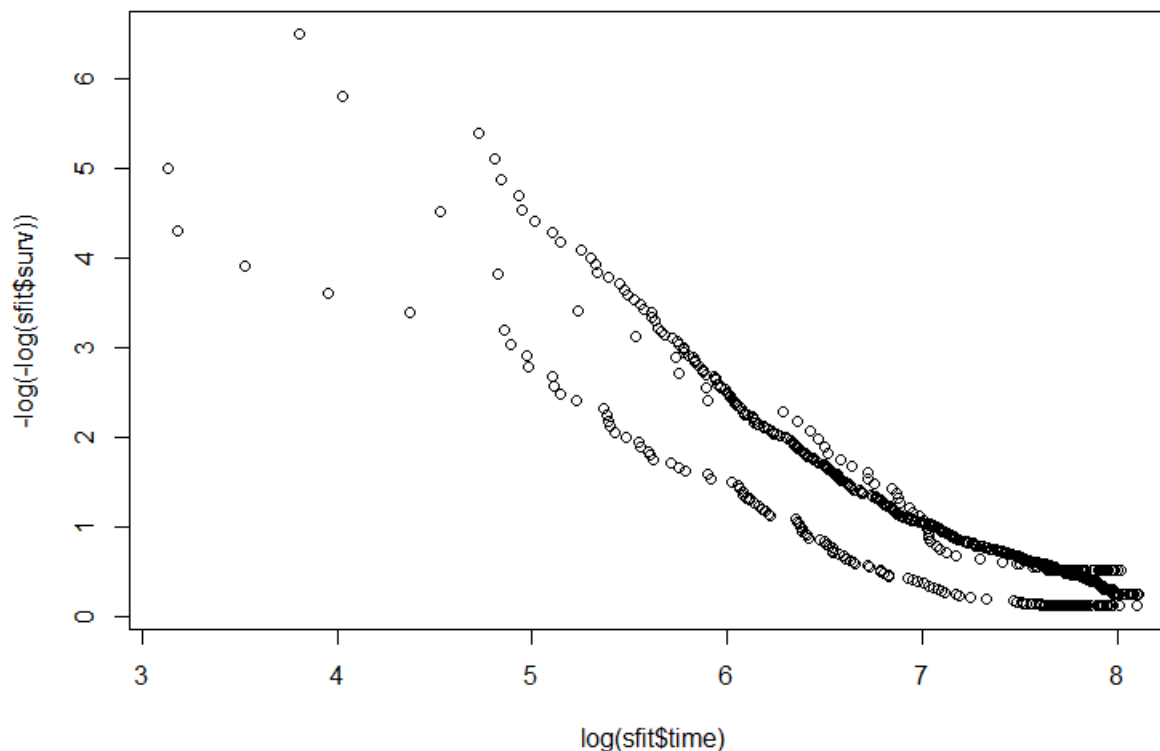
1. Túlélési görbék vizsgálata szemmel



```
> sfit <- survfit( formula = Surv(time, status) ~ differ, data = colon.death )  
> plot( sfit, mark.time = T, conf.int = F, xlab = "Time [day]", ylab = "Probability of survival" )
```

Arányos kockázat fennállásának tesztelése

2. log-kumulatív hazard görbék vizsgálata szemmel
 $\log(\text{time})$ vs. $-\log(-\log(S(t)))$



```
> sfit <- survfit( formula = Surv(time, status) ~ differ, data = colon.death )  
> plot( log(sfit$time), -log(-log(sfit$surv)) )
```

Arányos kockázat fennállásának tesztelése

3. Schoenfeld reziduálisok tesztelése

Az események bekövetkezésekor a kovariánsok megfigyelt és várható értékének különbsége.

```
> fit <- coxph( formula = Surv(time, status) ~ factor(differ), data = colon.death )  
> temp <- cox.zph(fit)  
> print(temp)
```

	rho	chisq	p
factor(differ)2	0.00955	0.0402	8.41e-01
factor(differ)3	-0.12644	6.9913	8.19e-03
GLOBAL	NA	19.4294	6.04e-05

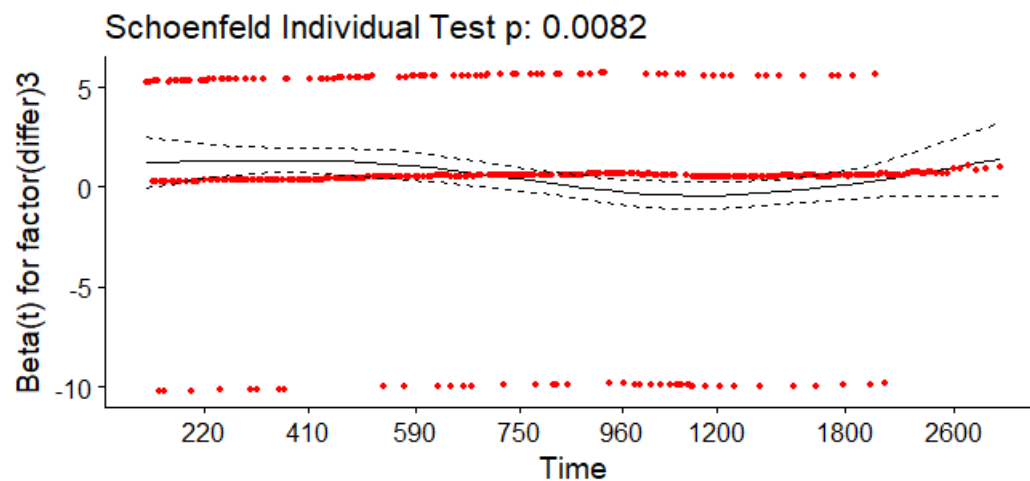
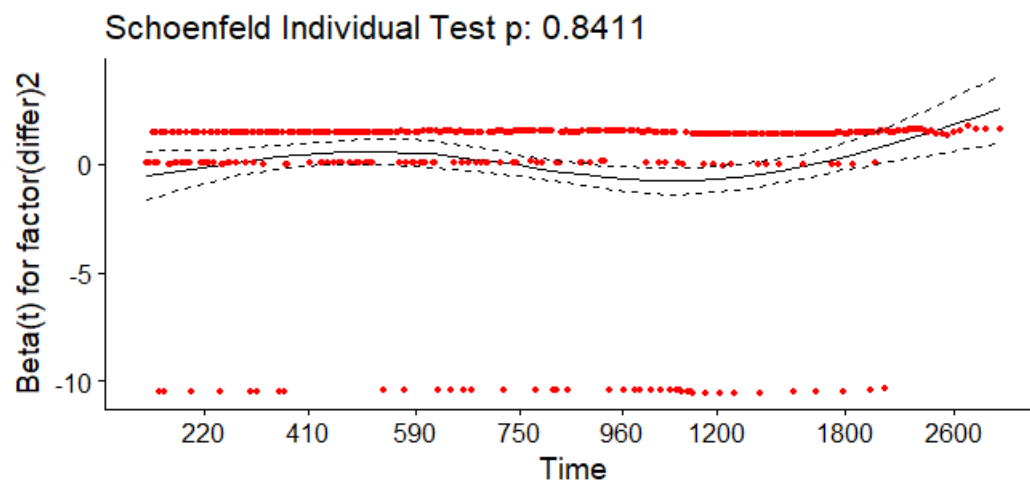
P < 0.05 esetén nem teljesül az arányos kockázat feltétele!

Arányos kockázat fennállásának tesztelése

Global Schoenfeld Test p: 6.039e-05

3. Schoenfeld reziduálisok tesztelése/ábrázolása

```
> library(survminer)
> ggcoxzph(temp)
```



Arányos kockázat fennállásának tesztelése a többváltozós modellen

```
> fit <- coxph( formula = Surv(time, status) ~ rx + sex + age + perfor + adhere,  
+ data = colon.death )  
> temp <- cox.zph( fit )  
> print(temp)
```

	rho	chisq	p
rxLev	-0.0446	0.897	0.344
rxLev+5FU	-0.0527	1.261	0.261
sex	0.0459	0.943	0.331
age	-0.0263	0.349	0.554
perfor	0.0348	0.557	0.456
adhere	0.0196	0.177	0.674
GLOBAL	NA	3.670	0.721

Feladatok

További változók vizsgálata Cox-regressziós modellel

Az arányos kockázat fennállásának tesztelése szemmel és a Schoenfeld reziduálisok kiszámításával