# Túlélés analízis

**Probléma:**

- Túlélési idő vizsgálata speciális vizsgálati módszereket igényel (pl. két csoport között az idők átlagait nem lehet direkt módon összehasonlítani)
- A túlélési idő nem normális eloszlású (korlátozó feltétel).

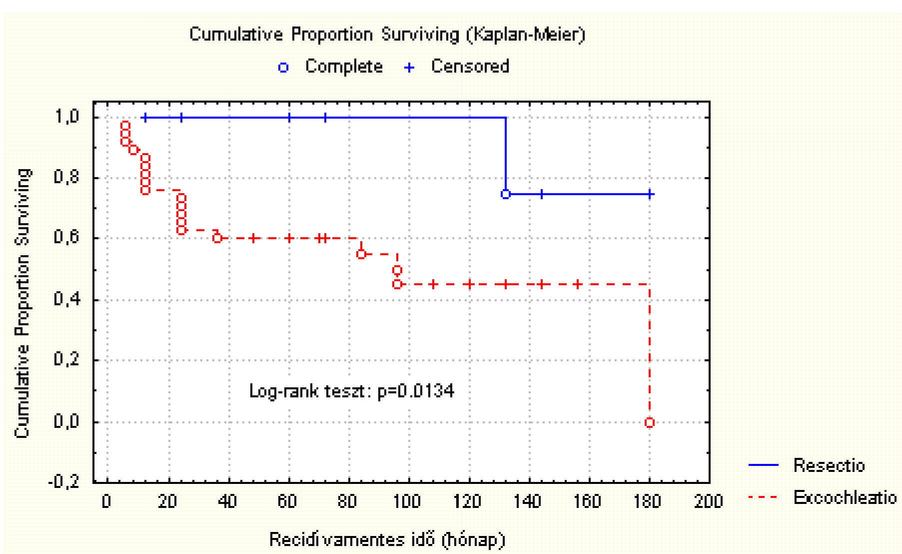
**Történeti háttér:**

- Először életbiztosítási statisztikusok (actuarial) használták (Berkson és Gage 1950).
- Cutler és Ederer fejlesztette tovább az eljárást

**Definíciók:**

- *Esemény (endpoint, failure):* bármilyen megfigyelés eredménye (műtét utáni felépülés).
- *Megfigyelési idő (observation time):* Az esetek többségében egy előre definiált időintervallum.
- *Követés (follow-up):* A betegek sorsának figyelése.
- *Kiesett egyének (drop-out):* a vizsgálatból kiesett egyének.
- *Visszavonás (withdrawing):* pl. együttműködés hiánya miatt egyéneket kizárunk a vizsgálatból.
- *Túlélési idő (survival time):* a megfigyelés kezdetétől a vizsgált eseményig eltelt idő (pl. hány nap tellik el az operációt követően a felépülésig).
- *Nem teljes adat (censored data):* olyan egyének, akik 'elvesznek' a megfigyelés során (pl. elköltöznek), de addigi adataikat felhasználjuk.

Az elemzés egyik általános eszköze a túlélési függvény megrajzolása (jellegzetes lépcsős függvény): A grafikon csont tumor esetén alkalmazott két műtéti eljárás eredményét jeleníti meg az idő függvényében az összehasonlító statisztikai eljárás eredményével együtt:



A görbe a 0. időponttól indul jelezvén a megfigyelés kezdetét. Ebben az időpontban mindenki komplikációmentes, így a függvény maximális értékű, azaz 1 (vagy 100%). Valahányszor egy esemény bekövetkezik, a függvényértéke csökken. A ° jel jelzi a grafikonon az eseményt. (A számítógépes programok különböző módon jelzik az eseményeket.) A görbén a $\perp$ jel jelzi, hogy nem következett be a vizsgált esemény (cenzorált) . A programok általában ezt a jelölést használják ilyen célra.

Kétfajta módon készíthetünk túlélési függvényt:

- Kaplan–Meier eljárás (product–limit módszer)
- life–table analízis (az előző módszert is tartalmazza).

## 1. **Kaplan–Meier becslés**

Tekintsük a következő hipotetikus vizsgálatot: 10 személy sorsát követjük a 10 napos megfigyelés alatt. Az esemény (end point) a bekövetkezett halál. A vizsgálat során 6 fő meghalt, 3 adata cenzorált, 1 fő a megfigyelés végén még élt.

A számoláshoz a következő adatokra lesz szükségünk:

- $t_i$: a vizsgált megfigyelési időpont
- r: a $t_i$–nél bekövetkezett halálesemény száma
- n: az élők száma a $t_i$ időpontnál
- $p_a$: a $t_i$ időpontnál élők aránya (n – r) / n
- p: a $t_i$ időponton túli élés becsült valószínűsége, ami a $t_i$–nél számított $p_a$ és az előtte ($t_{i-1}$ –nél) lévő $p_{i-1}$ szorzata
- SE: a p érték standard hibája = $p\sqrt{(1-p)/n}$

Az egyes megfigyelési időpontokra vonatkozó számítások:

| | |
|---|---|
| t = 2 | az esemény cenzorált, nincs teendő |
| t = 3 | eddig az időpontig 9 egyén él (hiszen t = 2–ben egy egyén kiesett, n = 9), itt 1 fő meghal (r = 1), tehát $p_a$ = (n – r) / n = (9 – 1) / 9 = 8 / 9 = 0.89 és p = 1 • 0.89, az SE = $p\sqrt{(1-p)/n} = 0.89\sqrt{(1-0.89)/9} = 0.10$ |
| t = 4 | az esemény cenzorált, nincs teendő |
| t = 5 | idáig 7 egyén él (n = 7), amelyből itt 1 fő meghal (r = 1), tehát $p_a$ = (7 – 1) / 7 = 0.86, a p = 0.89 • 0.86 = 0.77 és SE = $0.77\sqrt{(1-0.77)/7} = 0.14$ |
| t = 7 | eddig 6 fő él (n = 6), itt 1 fő meghal (r=1), vagyis $p_a$ = (6 – 1) / 6 = 0.83, a p = 0.77 • 0.83 = 0.64 és SE = $0.64\sqrt{(1-0.64)/6} = 0.16$ |
| t = 8 | eddig 5 fő él (n = 5) itt 2 fő meghal (r = 2), vagyis vagyis $p_a$ = (5 – 2) / 5 = 0.6, a p = 0.64 • |

$0.6 = 0.38$ és SE = $0.38\sqrt{(1-0.38)/5} = 0.13$

t = 9 eddig 2 fő él (n = 2), itt 1 fő meghal (r = 1), vagyis $p_a = (2 - 1) / 2 = 0.5$, a $p = 0.38 \bullet 0.5 = 0.19$ és SE = $0.19\sqrt{(1-0.19)/2} = 0.12$

t = 10 az esemény cenzorált (az egyén a vizsgálat lezártakor él, a további sorsát nem tudjuk).

Foglaljuk táblázatba a számítás eredményeit

| $t_i$ | r | n | $P_a$ | p | SE |
|---|---|---|---|---|---|
| 0 | 0 | 10 | 1.0 | 1.0 | – |
| 3 | 1 | 9 | 0.89 | 0.89 | 0.10 |
| 5 | 1 | 7 | 0.86 | 0.77 | 0.14 |
| 7 | 1 | 6 | 0.83 | 0.64 | 0.16 |
| 8 | 2 | 5 | 0.60 | 0.38 | 0.13 |
| 9 | 1 | 2 | 0.50 | 0.19 | 0.12 |

Kaplan–Meier eljárás számítási táblázata

A p érték alapján a kumulatív túlélési grafikonja az idő függvényében megrajzolható.

**Túlélési görbék összehasonlítása:**

a) *Általánosított (generalizált) Wilcoxon teszt:* a korai események nagyobb súlyt kapnak.
b) *Mantel–Cox teszt:* egy exponenciális score teszt.. Lényegében egy log–rank teszt.
c) *Breslow–teszt:* a korai megfigyeléseket súlyozza, kevésbé érzékeny a későbbi eseményekre.
d) *Tarone–Ware teszt:* a Breslow és a Mantel–Cox teszt között helyezkedik el. Az eseményeket közepesen súlyozza.
e) *Peto–Prentice teszt:* az általánosított Wilcoxon teszttel analóg.
f) *Log-rank teszt:* a két csoport közötti mortalitási arány egymáskonstanszorosa.

**Vizsgálati meggondolások:**

1) A vizsgálat milyen mintaszámmal, milyen hosszú ideig tartson,mi a vizsgált esemény (end point).
2) Hogyan kezeljük a drop out eseteket.
3) Ismétlődő jelenségre ne végezzünk túlélés analízist.

Dr. Dinya Elek

4) A túlélési kritérium, a feltétel rendszer nem változhat meg a vizsgálat folyamán.

## 2. **Cox–regresszió**

Cox (1972) a túlélési problémák vizsgálatára dolgozta ki a proportional hazards regressziós módszerét (egyszerűen Cox modell). A vizsgált esemény kockázata a hazard rate (a kumulatív túlélési görbe meredeksége egy időintervallumban). A kockázatot úgyis definiálhatjuk, mint adott t időpontban a halál bekövetkezésének valószínűségét, amikor tudjuk, hogy az egyén a t idő előtt még él. A megfogalmazást kérdés formájában is feltehetjük: a beteg milyen valószínűséggel éli meg az öt évet a beavatkozás után, ha már három évet túlélt?
Ha az egyik csoportban a halál kockázata háromszorosa a másik csoporténak, akkor a kockázat állandó marad az egész vizsgálat folyamán. Ezt úgy mondjuk, hogy a két csoport hazard függvénye egymással arányos, proporcionális. Az előbbi példa kapcsán ez azt jelenti, hogy az első csoportban a halálesemény valószínűsége háromszor nagyobb, mint a másik csoportban.

A Cox–regressziós modell a vizsgált magyarázó változók relatív kockázatát becsli. A kapcsolatot mindig egy kockázatmentes (rizikófaktor mentes) csoporthoz, a baseline csoporthoz viszonyítjuk.

Jelölje h(t) a kockázat mértékét a t időpontban, $\underline{x} = [x_1, x_2, x_3, ..., x_n]$ a vizsgált magyarázó változók vektorát. A Cox–regressziós modell a kockázat nagyságát írja le a t időpontban, az $\underline{x}$ vektor értékei mellett

$$\log h(t, \underline{x}) = \log [h_0(t)] + \sum_{i=1}^{n} \beta_i x_i$$

Az egyenletben a $h_0(t)$ függvény az egyén kockázatát jelenti a t időpontban, amikor a kovariánsok mindegyike 0. Az egyenlet az alábbi formában is írható

$$p(T > t, \underline{x}) = \left[P_0(T > t)\right]^{\exp\left(\sum_{i=1}^{n} \beta_i x_i\right)}$$

Az egyenlet a t–nél nagyobb időtartam (T) túlélési valószínűségét adja meg.

A regressziós vizsgálat legfontosabb szempontja a $\beta_i$ regressziós együtthatók meghatározása. A modell lényegében logisztikus modellnek is tekinthető.
Tekintsük az alábbi példát:

Követéses vizsgálat során (amelynek időtartama 22 év volt), 30561 fős mintán (40 év felettiek) a tüdőrák rizikófaktorainak hatását tanulmányoztuk. A Cox–regresszió alkalmazásával az egyes rizikótényezők relatív kockázatát kívántuk meghatározni. A táblázat a férfiakra vonatkozó vizsgálat egy részeredményét tartalmazza:

| Rizikó–faktorok | Regressziós koefficiens ($\beta_i$) | Standard error (SE) | $\beta_i$ – k 95%–os CI–je ($\beta_A$; $\beta_F$) | p | Relatív kockázat (RR) | RR 95%–os CI–je ($RR_A$; $RR_F$) |
|---|---|---|---|---|---|---|
| Mérsékelten dohányos | 1.1138 | 0.3343 | (0.46, 1.77) | 0.001 | 3.1 | (1.6, 5.9) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Erős dohányos | 2.3326 | 0.2847 | (1.77, 2.89) | 0.001 | 10.3 | (5.9, 18.0) |
| Tüdő felületen hegesedés | 0.7286 | 0.2165 | (0.30, 1.15) | 0.001 | 2.1 | (1.4, 3.2) |
| A megfigyelt kora | 1.7281 | 0.2986 | (1.14, 2.31) | 0.001 | 5.6 | (3.1, 10.1) |

Férfiak tüdőrák kockázatának Cox–modellje

A rizikófaktorok mellett a modell által számolt regressziós koefficiens, annak standard hibája, a regressziós együttható 95%–os konfidencia intervalluma szerepel:

$$\text{alsó érték} = \beta_A = \beta_i - 1.96 \bullet SE$$

$$\text{felső érték} = \beta_F = \beta_i + 1.96 \bullet SE$$

A p a $\beta_i$ együttható 0–tól való szignifikáns eltérését jelenti (a $H_0$: $\beta_i = 0$).

Erős dohányos férfi esetén ez 10–szeres értéket jelent. A relatív kockázatot a $\beta_i$ értéke alapján számoljuk: $RR = e^{\beta i}$

Az utolsó oszlop a relatív kockázat 95%–os konfidencia intervallumát tartalmazza:

$$RR_A = e^{\beta_A}$$

$$RR_F = e^{\beta_F}$$

**Megfontolások:**

a) Érdemes a modellvizsgálat előtt a változók függetlenségvizsgálatát elvégezni.
b) A modell feltételezi a kockázat arányának időbeli állandóságát.
c) A mintaszám megválasztásánál alkalmazzunk az ökölszabályt, hogy minden kovariánsra legalább 5 esemény jusson.
d) A Cox–modellt gyakran alkalmazzák az exploratív vizsgálatok során hipotézisek felállítására.