

# 1. Analýza přežití

## 1.1. Základní pojmy

Analýza přežití je obor statistiky zabývající se popisem a analýzou dat, která korespondují době od vstupní události (tzv. čas počátku) do výskytu sledované události (tzv. koncový bod). Za **vstupní událost** můžeme pokládat například

- narození,
  - počátek léčby,
  - začátek nemoci,
  - vstup jedince do studie,
  - svatbu,
  - zavedení nového přístroje do výroby
- a jiné.

**Koncovou událostí** může být

- úmrtí jedince,
  - návrat příznaků nemoci,
  - uzdravení pacienta,
  - rozvod,
  - porucha přístroje
- a další.

Dobu mezi těmito dvěma událostmi označujeme jako **dobu přežití**.

Analýza přežití, jak vyplývá z předchozího, má velmi široké uplatnění, např.

- ve zdravotnictví,
- v průmyslu,
- v zemědělství, ...

Pro jednoduchost budeme za čas počátku považovat vstup jedince do nějaké studie či experimentu a za koncový bod smrt jedince.

## 1.2. Specifika dat v analýze přežití

Data analýza přežití nejsou vhodná ke zpracování standardními statistickými metodami.

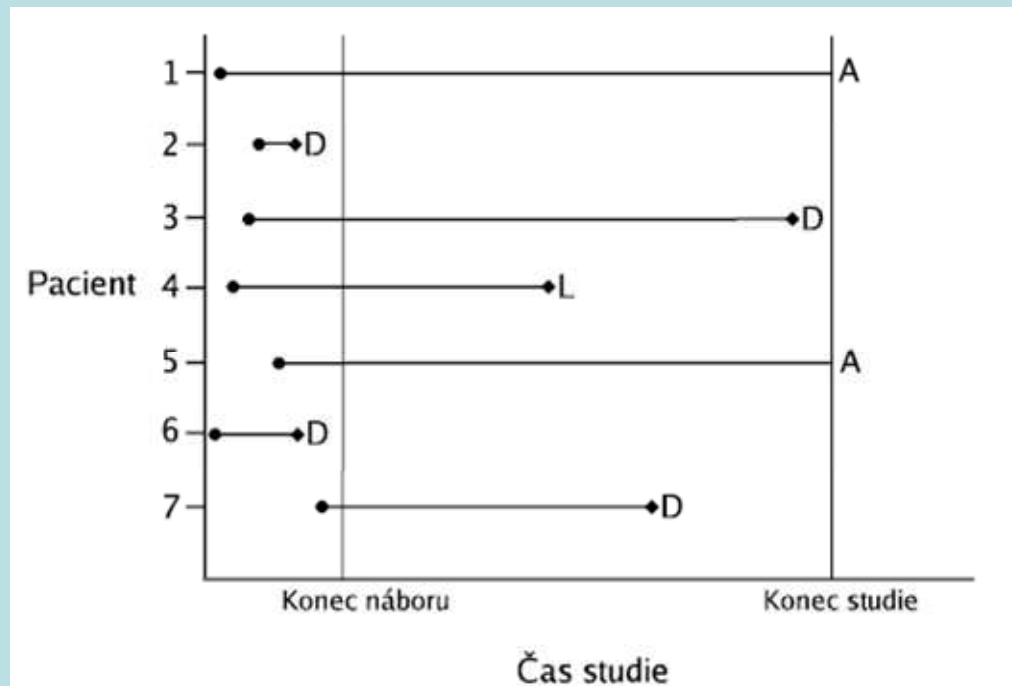
Důvody: doby přežití nejsou rozloženy symetricky a jsou často cenzorovány.

Doba přežití jedince je **cenzorována**, jestliže sledovaná koncová událost není u tohoto jedince během pozorování uskutečněna. To nastane například v případě, že

- s pozorovaným jedincem ztratíme kontakt, přestěhuje se nebo přestane docházet na pravidelné prohlídky nutné ke studii (nevíme, zda je na konci studie živ či mrtev);
  - data jsou zpracovávána v době, kdy se sledovaná událost u jedince ještě nevyskytla;
  - pozorovaný jedinec zemřel na jinou nemoc
- a podobně.

## 1.3. Typy cenzorování

- **cen-zorování zprava**: skutečná doba přežití jedince je delší než doba studie - nejčastější typ cen-zorování;
- **cen-zorování zleva**: jedinec zemře ještě před započítáním studie, například během výběru jedinců vhodných ke sledování;
- **intervalové cen-zorování**: odpovídá případu, že jedince je možno sledovat jen v určitých okamžicích (například jednou za měsíc), ne tedy bez přerušení po celou dobu trvání studie. V takovém případě vždy dostaneme jen informaci, že smrt nastala v časovém rozmezí určeném okamžikem posledního pozorování a současností.



**Obrázek: Ilustrace různých druhů cenzorování**

Na obrázku je znázorněna doba přežití u 7 pacientů. Období sledování je započato koncem náboru jedinců a ukončeno koncem studie. Písmeno D označuje smrt, L ztrátu kontaktu s jedincem a A znamená, že pacient je stále naživu.

**Upozornění:** Nadále budeme předpokládat, že data jsou cenzorovaná zprava, jelikož v praxi se jedná o nejčastěji se vyskytující typ cenzorování.

## 2. Funkcionální charakteristiky doby přežití a jejich odhady

### 2.1. Definice jednotlivých funkcí

Nechť spojitá nezáporná náhodná veličina  $T$  udává čas, který uplyne od počátku sledování jedince do jeho smrti. Rozložení pravděpodobností této náhodné veličiny můžeme popsat několika funkcemi.

**Distribuční funkce**  $F(t)$  je dána vztahem:

$$\forall t \geq 0 : F(t) = P(T \leq t)$$

$F(t)$  udává pravděpodobnost, že doba přežití jedince je nejvýše  $t$ .

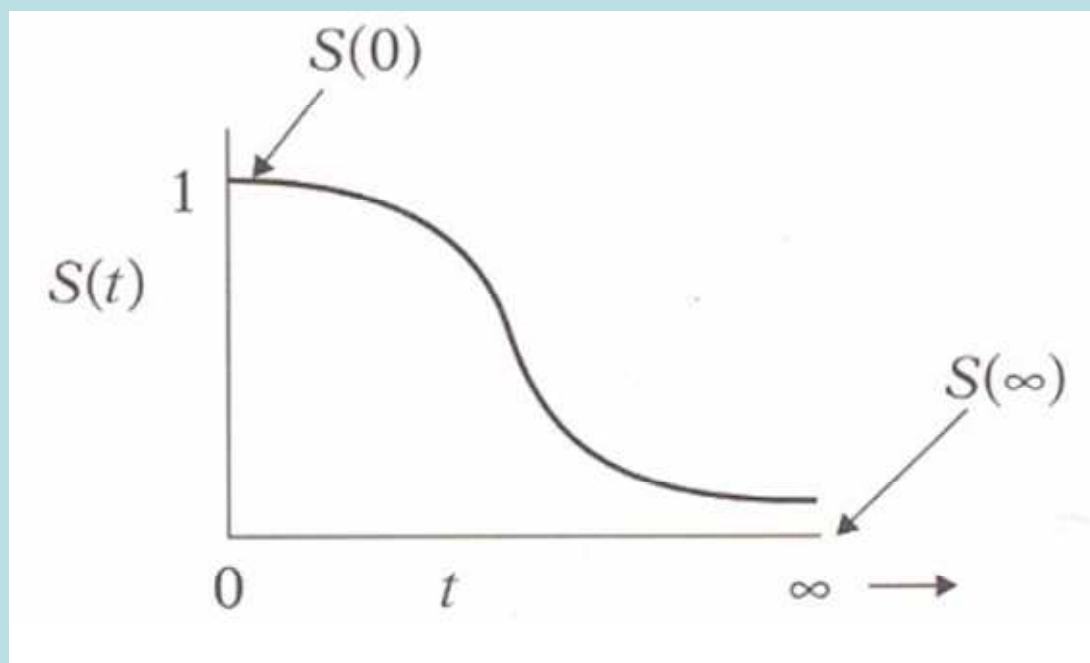
**Hustota pravděpodobnosti**  $f(t)$  je po částech spojitá nezáporná funkce, která je normovaná (tj. integrál z ní přes celý definiční obor je 1) a s distribuční funkcí je spjata integrálním vztahem:

$$\forall t \geq 0 : F(t) = \int_0^t f(u) du \text{ neboli } f(t) = \frac{dF(t)}{dt} \text{ ve všech bodech spojitosti funkce } f(t).$$

**Funkce přežití**  $S(t)$  je doplněk distribuční funkce do 1:

$$\forall t \geq 0 : S(t) = P(T > t)$$

$S(t)$  udává pravděpodobnost, že v čase  $t$  bude sledovaný jedinec ještě naživu.



**Obrázek: Teoretický tvar funkce přežití**

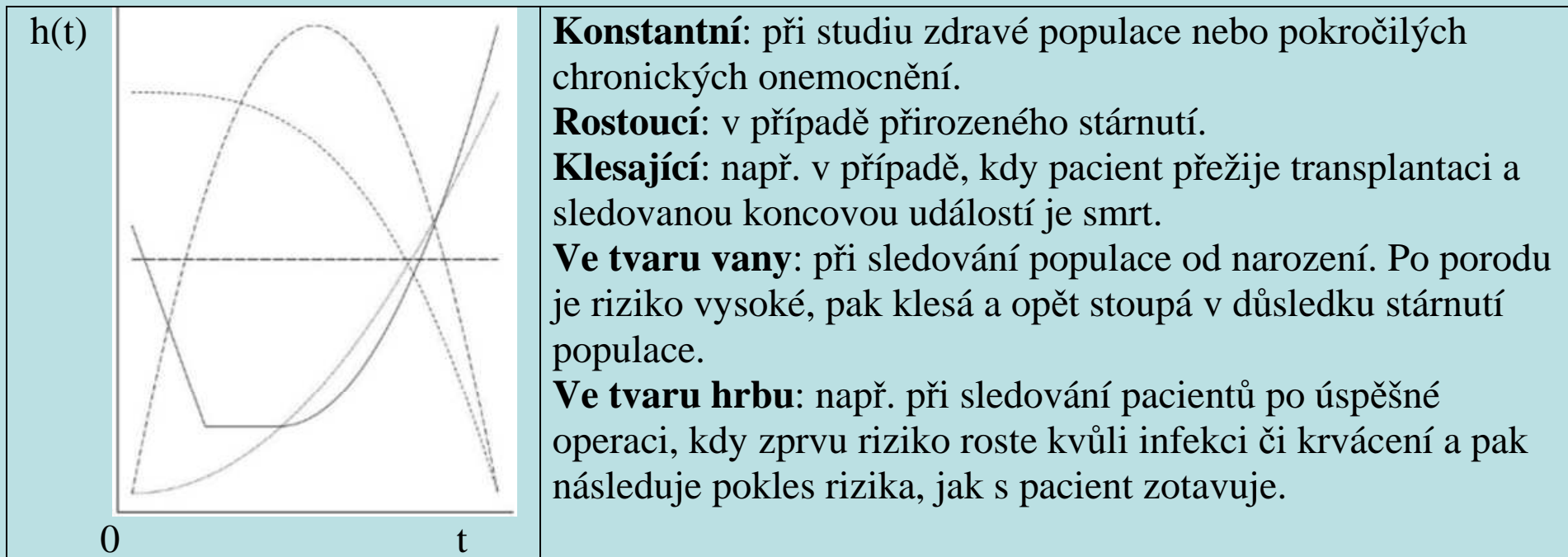
Vlastnosti funkce přežití:

- $S(t)$  je nerostoucí funkce
- v čase  $t = 0$  je  $S(t) = S(0) = 1$ , tj. na začátku studie u žádného jedince nenastala koncová událost, proto je pravděpodobnost přežití času  $t = 0$  rovna 1
- v čase  $t = \infty$  je  $S(t) = S(\infty) = 0$ , tj. teoreticky, překročí-li doba sledování určitý limit, žádný jedinec nebude přežívat a funkce přežití klesne k nule.

**Riziková funkce**  $h(t)$  je definována vztahem:

$$\forall t \geq 0 : h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t / T > t)}{\Delta t}$$

Hodnota rizikové funkce v bodě  $t$  udává pravděpodobnost, že jedinec zemře v čase  $t$  za podmínky, že přežil do tohoto času. Jde o „okamžitou intenzitu“ koncové události během velmi krátkého intervalu  $\Delta t$ .



**Obrázek: Teoretické tvary rizikové funkce**



**Kumulativní riziková funkce**  $H(t)$  vznikne integrací rizikové funkce  $h(t)$ :

$$\forall t \geq 0: H(t) = \int_0^t h(u) du$$

$H(t)$  představuje kumulativní riziko nastání koncové události v průběhu času.

**Vztahy mezi funkcionálními charakteristikami doby přežití:**

	$F(t)$	$S(t)$	$f(t)$	$h(t)$	$H(t)$
$F(t)$	x	$1-S(t)$	$\int_0^t f(u) du$	$1 - e^{-\int_0^t h(u) du}$	$1 - e^{-H(t)}$
$S(t)$	$1-F(t)$	x	$\int_t^\infty f(u) du$	$e^{-\int_0^t h(u) du}$	$e^{-H(t)}$
$f(t)$	$F'(t)$	$-S'(t)$	x	$h(t)e^{-\int_0^t h(u) du}$	$H'(t)e^{-H(t)}$
$h(t)$	$\frac{F'(t)}{1-F(t)}$	$-\frac{S'(t)}{S(t)}$	$f(t) / \int_t^\infty f(u) du$	x	$H'(t)$
$H(t)$	$-\ln(1-F(t))$	$-\ln S(t)$	$\int_0^t \left[ \frac{f(u)}{\int_t^\infty f(v) dv} \right] du$	$\int_0^t h(u) du$	x

## 2.2. Neparametrické odhady funkcionálních charakteristik doby přežití

### 2.2.1. Kaplanův - Meierův odhad funkce přežití

Neparametrická metoda, která poskytuje odhad funkce přežití v každém okamžiku, ve kterém došlo k alespoň jedné sledované události (smrti).

$n$  ... počet sledovaných jedinců

$t_1, t_2, \dots, t_r$  ... časy přežití sledovaných jedinců (Některá z těchto pozorování mohou být zprava cenzorovaná, takže se zde může vyskytnout několik jedinců se stejnou dobou přežití.)

$r$  ... počet časů úmrtí mezi  $n$  sledovanými jedinci ( $r \leq n$ )

$t_1 < t_2 < \dots < t_r$  ...  $r$  uspořádaných časů úmrtí

$n_j$  ... počet jedinců, kteří jsou živí před časem  $t_j$  (počet jedinců v riziku v čase  $t_j$ ),  $j = 1, 2, \dots, r$

$d_j$  ... počet jedinců, kteří zemřou v čase  $t_j$ ,  $j = 1, 2, \dots, r$

$\frac{n_j - d_j}{n_j}$  ... odhad pravděpodobnosti přežití pro interval  $\langle t_j, t_{j+1} \rangle$

Nyní předpokládejme, že smrti jedinců nastávají ve stejném okamžiku nezávisle na sobě.

**Kaplanův -Meierův odhad funkce přežití** je dán vztahem:

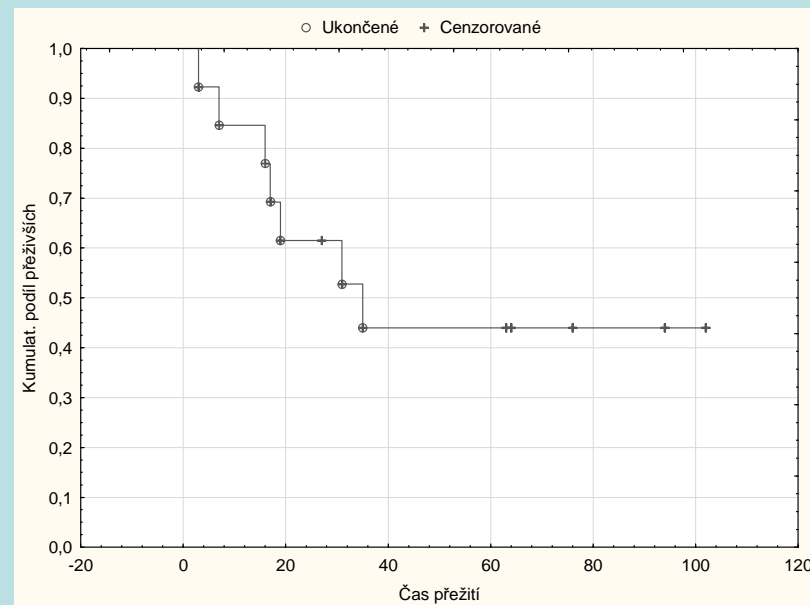
$$\hat{S}(t) = \begin{cases} 1 & \text{pro } t < t_1 \\ \prod_{j=1}^k \left( \frac{n_j - d_j}{n_j} \right) = \prod_{j=1}^k \left( 1 - \frac{d_j}{n_j} \right) & \text{pro } t_k \leq t < t_{k+1}, \quad k = 1, 2, \dots, r \text{ (klademe } t_{r+1} = \infty) \end{cases}$$

Kaplanův – Meierův odhad funkce přežití je dobře definovaný pro všechny doby přežití, které jsou menší než nejvyšší pozorovaný čas  $t_{\max}$ .

Jestliže největší pozorovaný čas je doba úmrtí, je  $\hat{S}(t) = 0$  pro  $t \geq t_r$ .

Jestliže však nejvyšší čas pozorování je cenzorovaný, hodnota funkce přežití je za tímto okamžikem neurčitá.

Graf funkce přežití odhadnuté K - M odhadem je nerostoucí, má schodovitý průběh s tím, že odhadnuté pravděpodobnosti přežití jsou konstantní mezi každými dvěma sousedními časy smrtí a v jednotlivých časech úmrtí funkce klesá.



**Obrázek: Kaplanův – Meierův odhad funkce přežití**

### 2.2.2. Intervaly spolehlivosti pro hodnoty funkce přežití

Rozptyl K – M odhadu funkce přežití je dán tzv. **Greenwoodovou formulí**:

$$D(\hat{S}(t)) \approx \hat{S}(t)^2 \sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)} \text{ pro } t_k \leq t \leq t_{k+1}, k = 1, 2, \dots, r \text{ (klademe } t_{r+1} = \infty)$$

100(1- $\alpha$ )% asymptotický **normální interval spolehlivosti** pro hodnotu S(t) funkce přežití v okamžiku t má meze:

$$d = \hat{S}(t) - \sqrt{D(\hat{S}(t))} \cdot u_{1-\alpha/2}, h = \hat{S}(t) + \sqrt{D(\hat{S}(t))} \cdot u_{1-\alpha/2}$$

Ze skutečnosti, že interval spolehlivosti je symetrický kolem  $\hat{S}(t)$ , vyplývá následující problém: je-li hodnota  $\hat{S}(t)$  blízká 0 nebo 1, může se stát, že  $d < 0$  či  $h > 1$ . V takovém případě se mezní hodnoty menší než 0 nahrazují nulou a větší než 1 jedničkou.

Pro malé rozsahy výběrů s velkým počtem cenzorovaných pozorování se doporučuje použít tzv. **log-log interval spolehlivosti**:

$$d = \hat{S}(t)e^{\Theta}, h = \hat{S}(t)e^{-\Theta}, \Theta = \sqrt{D(\ln(-\ln \hat{S}(t)))} \cdot u_{1-\alpha/2}, D(\ln(-\ln \hat{S}(t))) = \frac{\sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)}}{[\ln \hat{S}(t)]^2}$$

### 2.2.3. Kaplanův - Meierův odhad rizikové funkce

Rizikovou funkci  $h(t)$  na intervalu  $\langle t_j, t_{j+1} \rangle$  můžeme odhadnout pomocí vzorce:

$$\hat{h}(t) = \frac{d_j}{n_j(t_{j+1} - t_j)}, \quad t_j \leq t < t_{j+1}, \quad j = 1, 2, \dots, r$$

$d_j$  ... počet jedinců, kteří zemřou v čase  $t_j$ ,

$n_j$  ... počet jedinců, kteří jsou živí před časem  $t_j$ .

### 2.2.4. Kaplanův - Meierův odhad kumulativní rizikové funkce

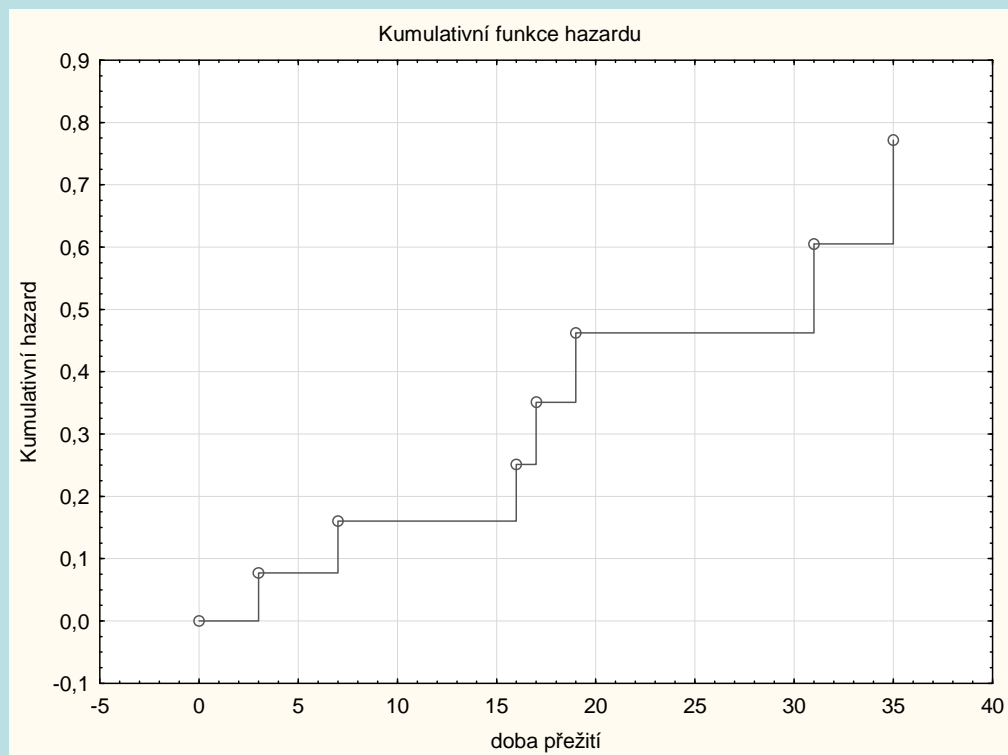
Na základě vztahu mezi funkcí přežití a kumulativní rizikovou funkcí lze získat odhad funkce  $H(t)$  ve tvaru:  $\hat{H}(t) = -\ln \hat{S}(t)$

### 2.2.5. Nelsonův – Aalenův odhad kumulativní rizikové funkce

$$\tilde{H}(t) = \begin{cases} 0 & \text{pro } t < t_1 \\ \sum_{j=1}^k \frac{d_j}{n_j} & \text{pro } t_k \leq t < t_{k+1}, \quad k = 1, 2, \dots, r \text{ (klademe } t_{r+1} = \infty) \end{cases}$$

Poznámka: Nelsonův – Aalenův odhad funkce přežití má tvar  $\tilde{S}(t) = e^{-\tilde{H}(t)}$ .

Graf kumulativní rizikové funkce je neklesající, má schodovitý průběh s tím, že odhadnuté pravděpodobnosti úmrtí jsou konstantní mezi každými dvěma sousedními časy smrtí a v jednotlivých časech úmrtí funkce roste.



**Obrázek: Kaplanův – Meierův odhad kumulativní rizikové funkce**

### **3. Srovnání dvou a více skupin jedinců podle doby přežití**

V analýze přežití se také zajímáme o to, zda existují statisticky významné rozdíly mezi různými skupinami, např. mezi pacienty s různými druhy léčby či mezi muži a ženami trpícími stejnou chorobou.

Nejjednodušší způsob je vykreslení odhadů funkce přežití do jednoho grafu. Výsledný graf již může být dostatečně informativní.

Existuje ovšem i řada statistických testů, které mohou být použity ke zjištění rozdílnosti mezi skupinami. Uvedeme log-rank test, Gehanův – Wilcoxonův test a chí-kvadrát test.

### 3.1. Příklad dvou skupin

Máme dvě skupiny jedinců. Předpokládáme, že v obou skupinách dohromady je  $r$  časů úmrtí,  $t_1 < t_2 < \dots < t_r$  a že v čase  $t_j$  zemře  $d_{1j}$  jedinců ze skupiny 1 a  $d_{2j}$  jedinců ze skupiny 2. Situace je ilustrována následující tabulkou:

Skupina	Počet úmrtí v čase $t_j$	Počet žijících po čase $t_j$	Počet v riziku před časem $t_j$
1	$d_{1j}$	$n_{1j} - d_{1j}$	$n_{1j}$
2	$d_{2j}$	$n_{2j} - d_{2j}$	$n_{2j}$
Celkem	$d_j$	$n_j - d_j$	$n_j$

Nulová hypotéza tvrdí, že neexistuje rozdíl v době přežití pro jedince z 1. a 2. skupiny, alternativní hypotéza tvrdí, že rozdíl existuje.

Jak log-rank test tak Gehanův – Wilcoxonův test jsou založeny na porovnání pozorovaného počtu úmrtí v 1. skupině  $d_{1j}$  a očekávaného (teoretického) počtu úmrtí v 1. skupině

$$e_{1j} = \frac{n_{1j}d_j}{n_j}, j = 1, 2, \dots, r.$$



### 3.1.1. Log-rank test

Zavedeme statistiku  $U_L = \sum_{j=1}^r (d_{1j} - e_{1j})$ . Její rozptyl  $D(U_L) = \sum_{j=1}^r \frac{n_{1j}n_{2j}d_j(n_j - d_j)}{n_j^2(n_j - 1)}$ .

Testová statistika  $W_L = \frac{U_L}{D(U_L)}$  se v případě platnosti nulové hypotézy asymptoticky řídí rozložením  $\chi^2(1)$ .

Nulovou hypotézu tedy zamítáme na asymptotické hladině významnosti  $\alpha$ , když  $W_L \geq \chi^2_{1-\alpha}(1)$ .

### 3.1.2. Gehanův – Wilcoxonův test

Zavedeme statistiku  $U_W = \sum_{j=1}^r n_j (d_{1j} - e_{1j})$ . Její rozptyl  $D(U_W) = \sum_{j=1}^r \frac{n_{1j}n_{2j}d_j(n_j - d_j)}{n_j - 1}$ .

Testová statistika  $W_W = \frac{U_W}{D(U_W)}$  se v případě platnosti nulové hypotézy asymptoticky řídí rozložením  $\chi^2(1)$ .

Nulovou hypotézu tedy zamítáme na asymptotické hladině významnosti  $\alpha$ , když  $W_W \geq \chi^2_{1-\alpha}(1)$ .

### 3.1.3. Srovnání Wilcoxonova a log-rank testu

Log-rank test přiřazuje relativně vyšší váhu pozdějším pozorováním než Gehanův - Wilcoxonův test. Pokud se ke konci studie vyskytuje hodně cenzorovaných dob přežití, doporučuje se použití Gehanova - Wilcoxonova testu.

Důležitým předpokladem log-rank testu je úměrnost rizikových funkcí obou skupin (předpoklad proporcionality rizik):  $h_1(t) = c h_2(t)$ ,  $c$  je konstanta nezávislá na čase.

Vizuální ověření: funkce přežití se nekříží.

Pokud se funkce přežití kříží, je vhodnější zvolit Gehanův - Wilcoxonův test.

## 3.2. Příklad tří a více skupin

Předpokládáme, že máme  $p \geq 3$  skupin jedinců. Nulová hypotéza tvrdí, že neexistuje rozdíl v době přežívání pro jedince z uvažovaných  $p$  skupin zatímco alternativa tvrdí, že aspoň mezi dvěma skupinami rozdíl existuje.

Tzv. **chí-kvadrát test** je založen na zobecnění log-rank testu nebo Gehanova - Wilcoxonova testu.

Zavedeme statistiky:

$$U_{Lk} = \sum_{j=1}^r \left( d_{kj} - \frac{n_{kj}d_j}{n_j} \right), \quad U_{Wk} = \sum_{j=1}^r n_j \left( d_{kj} - \frac{n_{kj}d_j}{n_j} \right), \quad k = 1, 2, \dots, p-1.$$

Tyto statistiky uspořádáme do sloupcových vektorů  $\mathbf{U}_L$  a  $\mathbf{U}_W$ , každý má  $p-1$  složek.

Kovariance mezi statistikami  $U_{Lk}$  a  $U_{Lk'}$  je dána vztahem

$$V_{Lkk'} = \sum_{j=1}^r \frac{n_{kj}d_j(n_j - d_j)}{n_j(n_j - 1)} \left( \delta_{kk'} - \frac{n_{k'j}}{n_j} \right) \text{ pro } k, k' = 1, 2, \dots, p-1, \text{ kde } \delta_{kk'} = \begin{cases} 1 \text{ pro } k = k' \\ 0 \text{ jinak} \end{cases}.$$

Tyto kovariance můžeme zapsat do varianční matice  $\mathbf{V}_L$  řádu  $p-1$ .

Analogicky, kovariance mezi statistikami  $U_{Wk}$  a  $U_{Wk'}$  je dána vztahem

$$V_{Wkk'} = \sum_{j=1}^r n_j^2 \frac{n_{kj}d_j(n_j - d_j)}{n_j(n_j - 1)} \left( \delta_{kk'} - \frac{n_{k'j}}{n_j} \right) \text{ pro } k, k' = 1, 2, \dots, p-1, \text{ kde } \delta_{kk'} = \begin{cases} 1 \text{ pro } k = k' \\ 0 \text{ jinak} \end{cases}.$$

Tyto kovariance můžeme zapsat do varianční matice  $\mathbf{V}_W$  řádu  $p-1$ .

Testová statistika má tvar  $\mathbf{U}_L^T \mathbf{V}_L^{-1} \mathbf{U}_L$  resp.  $\mathbf{U}_W^T \mathbf{V}_W^{-1} \mathbf{U}_W$ . V případě platnosti nulové hypotézy se testová statistika asymptoticky řídí rozložením  $\chi^2(p-1)$ .

Nulovou hypotézu tedy zamítáme na asymptotické hladině významnosti  $\alpha$ , když

$$\mathbf{U}_L^T \mathbf{V}_L^{-1} \mathbf{U}_L \geq \chi^2_{1-\alpha}(p-1) \text{ resp. } \mathbf{U}_W^T \mathbf{V}_W^{-1} \mathbf{U}_W \geq \chi^2_{1-\alpha}(p-1).$$

## Příklad

Datový soubor obsahuje informace o 67 pacientech s nádorem slinivky břišní, kteří byli operováni na chirurgické klinice fakultní nemocnice v Motole. Sběr dat proběhl v letech 1995 - 2005. Za sledovanou událost budeme považovat smrt pacienta. V době ukončení studie 15 pacientů žilo, 52 zemřelo na sledovanou diagnózu.

U každého pacienta bylo sledováno 6 proměnných.

Věk ... věk pacienta v letech

Sex ... pohlaví pacienta (1 – muž, 2, žena)

Doba sledování ... doba sledování pacienta v měsících

Smrt ... indikátor úmrtí pacienta (0 – žije, 1 – nežije)

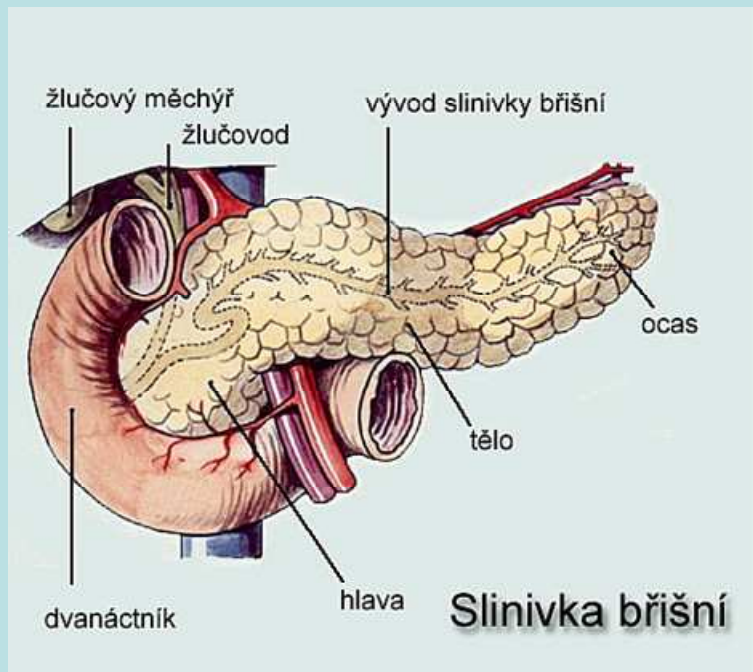
Stadium ... stadium rakoviny, 1 až 4

Typ ... typ nádoru (1 - duktální adenokarcinom, 2 - ampulární karcinom, 3 - karcinom terminálního choledochu)

Data jsou uložena v souboru slinivka.sta (resp. slinivka.xls).

## Slinivka břišní (pankreatis)

Je to žláza patřící k orgánům trávicí soustavy, je velká 12 – 16 cm a váží 60 – 90 g. Produkuje trávicí enzymy (exogenní funkce) a také hormony, nejdůležitější jsou inzulín a glukagon (endogenní funkce).



Skládá se z hlavy (caput pancreatis), těla (corpus pancreatis) a ocasu (cauda pancreatis).

## **Karcinom slinivky**

Je to zhoubný nádor postihující exokrinní část slinivky. Má vysokou mortalitu, 5 let od stanovení diagnózy se dožije jen 4 % nemocných, většina se nedožije ani jednoho roku. Onemocnění je současnými prostředky často diagnostikováno až v pokročilém stádiu. Klinicky je často dlouho bez symptomů.

Vnější rizikové faktory:

Věk (nejčastěji se vyskytuje v 7. a 8. dekádě života)

Kouření (zvyšuje riziko onemocnění 2x až 3x)

Alkohol (nepřímý rizikový faktor, může způsobit chronický zánět slinivky, což je jeden z nejvýraznějších vnitřních faktorů)

Složení potravy (jako rizikový faktor působí vysoký příjem masa, cholesterolu, smažených pokrmů)

Vnitřní rizikové faktory:

Chronický zánět slinivky (zvláště hereditární pankreatitida)

Cukrovka

Rodinná zátěž

Stav po odstranění žlučníku

Léčba

Chirurgická (kurativní nebo paliativní)

Radioterapie

Chemoterapie (obě léčby jsou jen paliativní, mají malou účinnost)

Prognóza tohoto onemocnění je velmi špatná. Šance, jak ji zlepšit nebo prodloužit život pacientovi, je zachytit nádorové bujení včas, což se pro asymptomaticnost stále nedaří.

Známé osobnosti, které zemřely na rakovinu slinivky:

Steve Jobs

Patrick Swayze

Pavel Dostál

Richard Hess

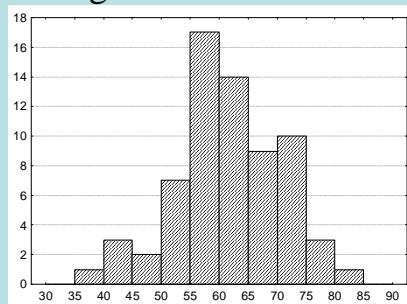


# Popis datového souboru

## Číselné charakteristiky věku

Proměnná	N platných	Průměr	Medián	Minimum	Maximum	Sm. odch.	Šikmost	Špičatost
věk	67	62,1	63	36	81	9,18	-0,36	0,20

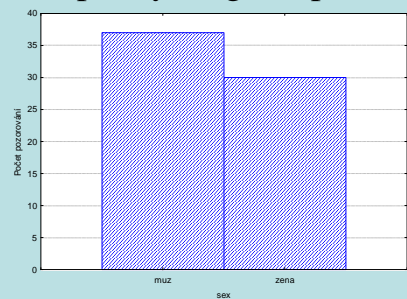
## Histogram věku



## Četnostní tabulka proměnné sex

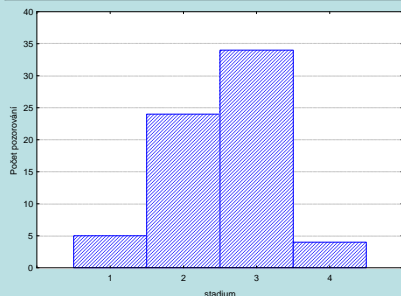
Kategorie	Četnost	Kumulativní četnost	Rel. četnost	Kumulativní rel. četnost
muz	37	37	55,22	55,22
zena	30	67	44,78	100,00

## Sloupkový diagram proměnné sex



## Četnostní tabulka proměnné stadium

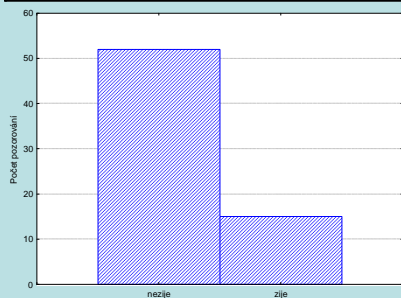
Tabulka četností:stadium (slinivka.sta)				
Kategorie	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
1	5	5	7,46	7,46
2	24	29	35,82	43,28
3	34	63	50,75	94,03
4	4	67	5,97	100,00



Vidíme, že nejčastěji se rakovinu slinivky podaří odhalit ve stadiu 3, jelikož u 37 jedinců z celkových 67 byla rakovina objevena právě v tomto stadiu, což je přibližně 50,7 %. U 24 jedinců byl karcinom zjištěn ve stadiu 2, tedy přibližně v 35,8% případů, ve stadiu 1 byl objeven u 5 pacientů, což je přibližně 7,5% a ve stadiu 4 u 4 jedinců, tudíž v 6% pozorování.

## Četnostní tabulka proměnné smrt

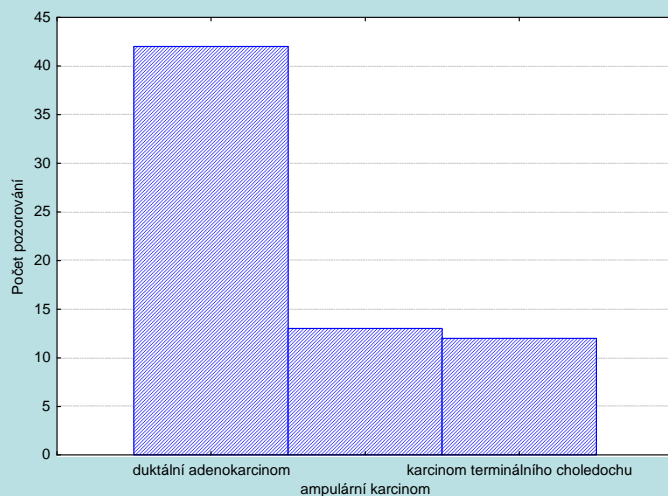
Kategorie	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
žije	15	15	22,39	22,39
nezije	52	67	77,61	100,00



Na karcinom umřelo 52 pacientů, tj. 77,6%, žije 15 pacientů, tj. 22,4%.

## Četnostní tabulka proměnné typ

Kategorie	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
duktální adenokarcinom	42	42	62,69	62,69
ampulární karcinom	13	55	19,40	82,09
karcinom terminálního choledochu	12	67	17,91	100,00



Nejvíce jedinců, 42, je postiženo duktálním adenokarcinomem, což je přibližně 62,7 % ze všech 67 pozorování, 13 pacientů má ampulární karcinom, těch je přibližně 19,4 %, a poslední typ karcinomu - karcinom terminálního choledochu - má 12 jedinců, což činí přibližně 17,9 % všech sledovaných jedinců.

Nyní se zaměříme na charakteristiky doby přežití pro jednotlivé skupiny pacientů. Způsob výpočtu ukážeme pro proměnnou typ.

Statistiky – Pokročilé lineární/nelineární modely – Analýza přežívání - Porovnání více vzorků – OK – Proměnné – Přežívání doba sledování, Cenzor. prom. smrt, Grupovací prom. typ – OK – Kódy pro ukončené nežije, Kódy pro cenzorované žije, Kódy skupin Vše – OK – OK – záložka Popisné statistiky – Popisné statistiky.

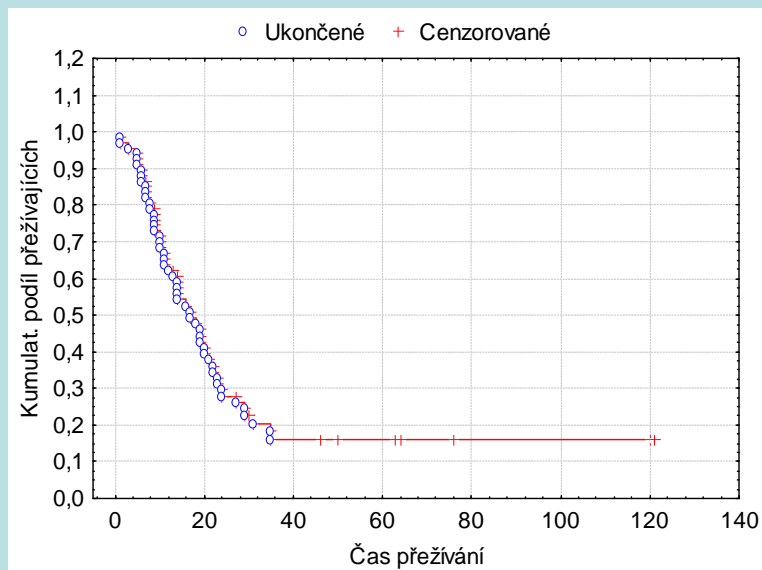
Skupina	Popisné statistiky pro každou skupinu (slinivka.sta)					
	Medián	Průměr	SmOdch	PčNecenz	PčCenzor	CelkPčt
duktální adenokarcinom	14,0	16,0	8,3	37	5	42
ampulární karcinom	27,0	36,8	34,7	8	5	13
karcinom terminálního choledochu	9,5	26,0	33,8	7	5	12
Celkem	14,0	21,9	22,8	52	15	67

Pacienti s typem karcinomu duktální adenokarcinom se v průměru dožívají přibližně 16 měsíců, s ampulárním karcinodem 36,8 měsíců a jedinci, u kterých se vyskytl karcinom terminálního choledochu, se v průměru dožívají 26 měsíců.

Výsledky pro proměnnou sex

Skupina	Popisné statistiky pro každou skupinu (slinivka.sta)					
	Medián	Průměr	SmOdch	PčNecenz	PčCenzor	CelkPčt
muz	19,0	21,4	14,6	30	7	37
zena	12,5	22,5	30,3	22	8	30
Celkem	14,0	21,9	22,8	52	15	67

Nyní získáme Kaplanův – Meierův odhad funkce přežití pro celý soubor:  
Statistiky – Pokročilé lineární/nelineární modely – Analýza přežívání – Kaplan – Meierova metoda - Proměnné – Přežívání doba sledování, Indikátor cenzorov. smrt – OK – Kódy pro ukončené nežije, Kódy pro cenzorované žije, Kódy skupin Vše – OK – OK – záložka K – M grafy – Časy přežívání vs. Kum. Podíly přežívajících



Jak je patrné z grafu, velká část pacientů s tímto druhem karcinomu umírá brzy po zjištění nádoru, jelikož odhad funkce přežití strmě klesá.

Dále vypočteme vybrané kvantily odhadnuté funkce přežití:

Statistiky – Pokročilé lineární/nelineární modely – Analýza přežívání – Kaplan – Meierova metoda - Proměnné – Přežívání doba sledování, Cenzor. prom. smrt – OK – Kódy pro ukončené nežije, Kódy pro cenzorované žije, Kódy skupin Vše – OK – OK – záložka Detaily – Kvantily funkce přežívání

Kvantily	Kvantily (slinivka.sta) funkce přežívání
	Čas přeživ.
25. kvantil (dolní kvartil)	9,0
50. kvantil (medián)	17,0
75. kvantil (horní kvartil)	28,4

Z těchto údajů vyplývá, že jedna čtvrtina pacientů umírá do 9. měsíce od zjištění nádoru, polovina jedinců zemře do 17. měsíce, a přibližně do 28,4 měsíců zemřou tři čtvrtiny jedinců.

## Porovnání doby sledování pro muže a ženy

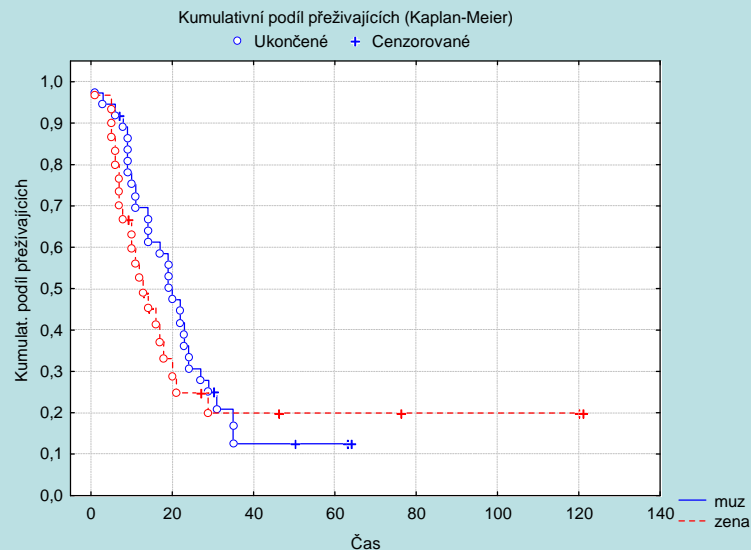
Mužů bylo celkově 37 a zemřelo jich 30, což je 81,1 % ze všech mužů, žen bylo 30 a umřelo jich 22, tedy 73,3 %. Tato situace je zachycena v kontingenční tabulce proměnných sex a smrt.

	Kontingenční tabulka (slinivka.sta)			Řádk. součty
	smrt	sex muz	sex zena	
Četnost	zije	7	8	15
Sloupc. četn.		18,92%	26,67%	
Četnost	nezije	30	22	52
Sloupc. četn.		81,08%	73,33%	
Četnost	Vš.skup.	37	30	67

Podíl šancí na přežití pro muže a ženy je  $OR = 0,641$ , tedy muži mají oproti ženám jen dvoutřetinovou šanci na přežití.

Do jednoho grafu nakreslíme odhadnuté funkce přežití pro muže a pro ženy:

Statistiky – Pokročilé lineární/nelineární modely – Analýza přežívání - Porovnání dvou vzorků – OK – Proměnné – Přežívání doba sledování, Cenzor. prom. smrt, Grupovací prom. sex – OK – Kódy pro ukončené nežije, Kódy pro cenzorované žije – OK – záložka Grafy funkcí – Kum. podíl přeživ. dle skupin (Kaplan Meier)



Vidíme, že funkce se během svého průběhu překříží a tudíž pro testování hypotézy, že se doby přežití v jednotlivých skupinách neliší, je výhodnější použít Gehanův - Wilcoxonův test. Budeme tedy testovat hypotézu  $H_0$ : doby přežití se pro jednotlivá pohlaví neliší. Hladinu významnosti zvolíme  $\alpha = 0,05$ .



Vrátíme se do tabulky Výsledky dvouvýběrových testů – záložka Dvouvýběrové testy – Gehanův – Wilcoxonův test. Zajímá nás záhlaví výstupní tabulky:

Gehanův Wilcoxonův test (slinivka.sta)  
WW = 236,00 Sčt = 91546, Prom = 22980,  
Test. statist. = 1,553528 p = ,12030

Hodnota testové statistiky je rovna 1,5535 a příslušná p-hodnota je 0,1203. Jelikož je p-hodnota větší než 0,05, tak hypotézu  $H_0$  nezamítáme na asymptotické hladině významnosti 0,05. Tudíž s rizikem nejvýše 5% můžeme říci, že doby přežití u mužů a žen s rakovinou slinivky nejsou rozdílné.

## Srovnání doby sledování u jednotlivých typů karcinomu

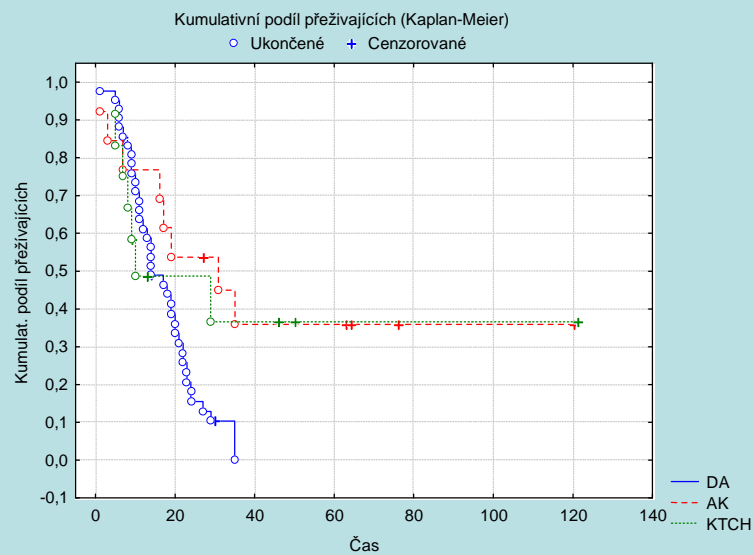
Pacientů s duktálním adenokarcinomem bylo 42, umřelo jich 37, tj. 88,1%.

Pacientů s ampulárním karcinomem bylo 13, umřelo jich 8, tj. 61,5%.

Pacientů s karcinomem terminálního choledochu bylo 12, umřelo jich 7, tj. 58,3%. Tato situace je zachycena v kontingenční tabulce proměnných smrt a typ.

Kontingenční tabulka (slinivka.sta)					
Tab. :					
	smrt	typ duktální adenokarcinom	typ ampulární karcinom	typ karcinom terminálního choledochu	Řádk. součty
Četnost	zije	5	5	5	15
Sloupc. četn.		11,90%	38,46%	41,67%	
Četnost	nezije	37	8	7	52
Sloupc. četn.		88,10%	61,54%	58,33%	
Četnost	Vš.skup.	42	13	12	67

Do jednoho grafu nakreslíme odhadnuté funkce přežití pro všechny tři skupiny pacientů:  
Statistiky – Pokročilé lineární/nelineární modely – Analýza přežívání - Porovnání více vzorků –  
OK – Proměnné – Přežívání doba sledování, Cenzor. prom. smrt, Grupovací prom. typ – OK –  
Kódy pro ukončené nežije, Kódy pro cenzorované žije – OK – záložka Grafy funkcí – Kumul.  
podíl přeživ. (Kaplan Meier) dle skupin.



K testování nulové hypotézy, že typ karcinomu nemá vliv na dobu přežití pacientů, použijeme chí-kvadrát test, hladinu významnosti zvolíme 0,05.

Vrátíme se do tabulky Výsledky porovnání přežívání ve více skupinách - Výpočet. Zajímá nás záhlaví výstupní tabulky:

Proměnné : přežití by typ (3 skupiny (slinivka.sta)

Cenzor. prom. : smrt (Cenzor. případy jsou značeny +)

Chi2 = 2,04370 sv= 2 p = ,35994

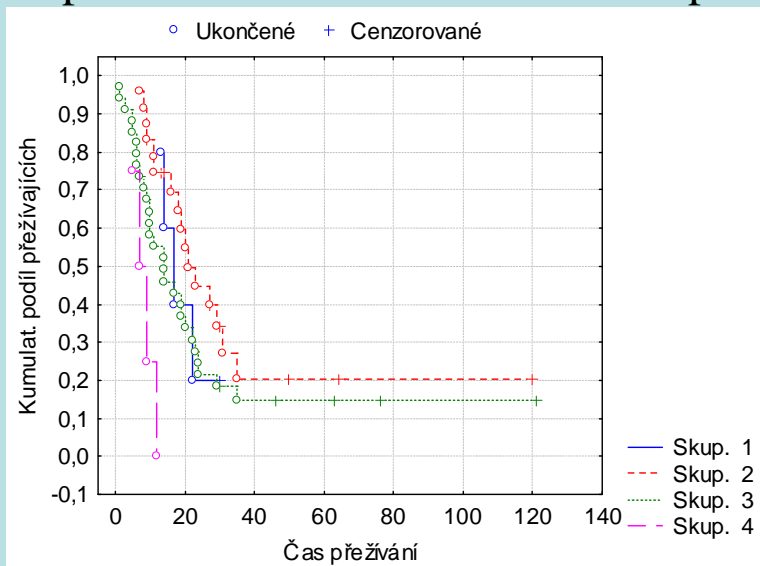
Hodnota testové statistiky je 2,0437 a příslušná p-hodnota je 0,3599. Poněvadž je p-hodnota větší než 0,05, tak nulovou hypotézu o shodě dob přežití u jednotlivých typů karcinomů nezamítáme na asymptotické hladině významnosti 0,05.

## Srovnání doby sledování u jednotlivých stadií rakoviny

Stadium 1 mělo 5 pacientů, tj. 7,5%. Stadium 2 mělo 24 pacientů, tj. 35,8 %. Stadium 3 mělo 34 pacientů, tj. 50,7 %. Stadium 4 měli 4 pacienti, tj. 6 %. Ze skupiny 5 pacientů v 1. stadiu nemoci umřeli 4 pacienti, tj. 80%. Ze skupiny 24 pacientů ve 2. stadiu nemoci umřelo 16 pacientů, tj. 66,7 %. Ze skupiny 34 pacientů ve 3. stadiu nemoci umřelo 28 pacientů, tj. 82,3 %. Ze skupiny 4 pacientů ve 4. stadiu nemoci umřeli všichni, tj. 100 %. Výsledky jsou přehledně zachyceny v kontingenční tabulce proměnných smrt a stadium.

Kontingenční tabulka (slinivka.sta)						
Tab. :						
	smrt	stadium 1	stadium 2	stadium 3	stadium 4	Řádk. součty
Četnost	zije	1	8	6	0	15
Sloupc. četn.		20,00%	33,33%	17,65%	0,00%	
Četnost	nezije	4	16	28	4	52
Sloupc. četn.		80,00%	66,67%	82,35%	100,00%	
Četnost	Vš.skup.	5	24	34	4	67

## Kaplanův – Meierův odhad funkce přežití pro čtyři skupiny pacientů rozlišených podle stadia



Testujeme hypotézu  $H_0$ : přežívání v daných čtyřech skupinách se neliší.

Hodnota testové statistiky chí-kvadrát testu: 9,2471, počet stupňů volnosti = 3, p-hodnota = 0,0262,  $H_0$  tedy zamítáme na hladině významnosti 0,05.

## Shrnutí statistického zpracování:

- Zkoumali jsme data 67 pacientů, většina z nich byla ve věku nad 60 let (55,2 %), lehce převažovali muži (55,2 %).
- Na nádor slinivky zemřelo 77,6 % pacientů, kteří průměrně přežívali 14,7 měsíce.
- Nejčastější typ nádoru byl duktální adenokarcinom (62,7 %), který má nejkratší průměrnou dobu přežití, a to 16 měsíců.
- Nejvíce pacientů s nacházelo ve 3. stadiu rakoviny (50,7 %).
- K-M odhad funkce přežití ukazuje, že přežití pacientů prudce klesá do 35. měsíce od stanovení diagnózy. Po tomto měsíci už nebylo žádné úmrtí u zbylých 7 pacientů.
- Na 5% hladině významnosti se neprokázal rozdíl v době přežívání mezi muži a ženami ani mezi jednotlivými typy nádoru.
- Na 5% hladině významnosti se prokázal rozdíl v době přežívání mezi pacienty s různými stádii rakoviny.