Analýza přežití

Motivace: 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, třeba
ve zdravotnictví,
v průmyslu,
v zemědělství,
v demografii
apod.
Pro jednoduchost budeme za čas počátku považovat vstup jedince do nějaké studie či experimentu a za koncový bod smrt jedince.

Specifika dat v analýze přežití

Data analýza přežití nejsou vhodná ke zpracování standardními statistickými metodami používanými v analýze dat. Hlavním důvodem je fakt, že doby přežití 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

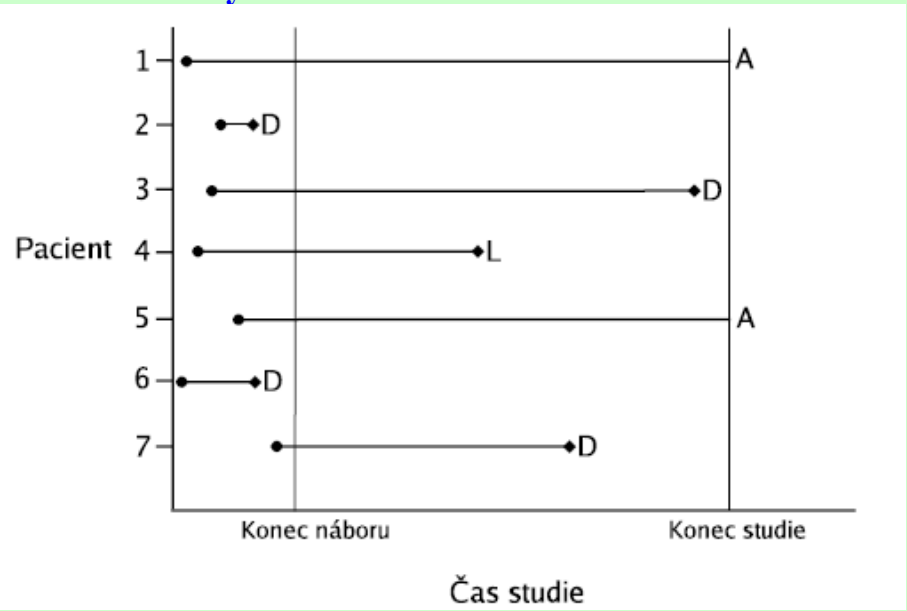
- s pozorovaným jedincem ztratíme kontakt, přestěhuje se nebo přestane docházet na pravidelné prohlídky nutné ke studii (nevíme, zda je na konci studie živ či mrtev)
- data jsou zpracovávána v době, kdy se sledovaná událost u jedince ještě nevyskytla
- pozorovaný jedinec zemřel na jinou nemoc
- a podobně.

V každé z těchto situací jedinec, který vstoupil do studie v čase t_0, zemřel v čase $t_0 + t$, avšak čas t je neznámý. Víme pouze, že jedinec byl živ v čase $t_0 + c$, kde čas c se nazývá cenzorovaná doba přežití. V tomto případě, kdy se cenzorované události staly napravo od posledního známého času přežití, mluvíme o cenzorování zprava - skutečná doba přežití je vyšší než doba pozorování.

V případě, že doba přežití jedince je menší než sledovaná, jedná se o další typ cenzorování, a to cenzorování zleva. Tímto druhem cenzorování je případ, kdy jedinec zemře dříve než oficiálně započne studie, například během výběru jedinců vhodných ke sledování.

Posledním typem cenzorování je intervalové cenzorování, které 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.

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. Na začátku studie zjistíme, že pacienti 2 a 6 již zemřeli, avšak nevíme přesně čas úmrtí. Jediné, co víme, je, že tito jedinci zemřeli někdy před začátkem studie, a tudíž se jedná o cenzorování zleva. V případě pacientů 1, 4 a 5 se jedná naopak o cenzorování zprava. Jelikož s pacientem 4 jsme během studie ztratili kontakt, neznáme jeho stav na konci studie. Pacienti 1 a 5 jsou na konci studie stále naživu, a tedy sledovaná událost, v našem případě smrt, se u nich během pozorování nevyskytla. Jedná se tedy také o cenzorovaný čas přežití zprava. Nakonec u jedinců 3 a 7 se sledovaná událost vyskytla během doby pozorování, a tudíž se jedná o necenzorované časy přežití, jelikož přesně víme dobu úmrtí.

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

Funkce přežití

Nechť spojitá nezáporná náhodná veličina T udává čas, který uplyne od počátku sledování jedince do jeho smrti. Rozložení pravděpodobností této náhodné veličiny je popsáno hustotou pravděpodobnosti $\varphi(t)$. Distribuční funkce $\Phi(t)$ je s hustotou spjata vztahem

$$\forall t \geq 0 : \Phi(t) = \int_0^t \varphi(u)\,du .$$

Zavedeme funkci přežití $\Psi(t) = P(T > t)$.
Hodnota funkce přežití v bodě t je pravděpodobnost, že doba přežití sledovaného jedince je větší než t. Funkci přežití lze pomocí hustoty vyjádřit vztahem

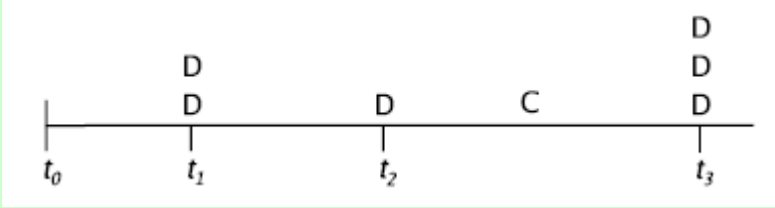
$$\forall t \geq 0 : \Psi(t) = \int_t^\infty \varphi(u)\,du$$

a pomocí distribuční funkce vztahem

$$\forall t \geq 0 : \Psi(t) = 1 - \Phi(t).$$

Kaplanův - Meierův odhad funkce přežití

Kaplanův - Meierův odhad funkce přežití je metoda, která poskytuje odhad funkce přežití v každém okamžiku, ve kterém došlo k alespoň jedné sledované události. Opět budeme pro jednoduchost za tuto událost považovat smrt.
K určení Kaplanova - Meierova odhadu funkce přežití z cenzorovaných dat se nejprve rozdělí doba pozorování do souboru časových intervalů. Každý z těchto intervalů je zkonstruován tak, aby v každém z nich bylo obsaženo alespoň jedno úmrtí, přičemž čas smrti je vzat jako počátek jednotlivých intervalů. Například předpokládejme, že t_1, t_2, t_3 jsou tři zaznamenané časy přežití uspořádané dle velikosti tak, že $t_1 < t_2 < t_3$, a c je cenzorovaný čas přežití, který spadá mezi časy t_2, t_3. Zkonstruované intervaly tedy začínají v časech t_1, t_2, t_3. Každý z intervalů obsahuje jeden čas úmrtí, ačkoliv zde může být více než jeden jedinec, který zemřel v některém z jednotlivých časů t_1, t_2, t_3. Stojí za povšimnutí, že žádný interval nezačíná v cenzorovaném čase c. Situace je ilustrována na následujícím obrázku, kde D reprezentuje smrt a C cenzorovaný čas přežití. Vidíme, že dva jedinci umřeli v čase t_1, jeden v čase t_2 a tři zemřeli v čase t_3.



Čas počátku, například studie, je označen jako t_0. Zde je také počátek prvního období, které končí před t_1, získáme tedy interval $\langle t_0, t_1 \rangle$. Tento interval neobsahuje žádné úmrtí. První zkonstruovaný interval $\langle t_1, t_2 \rangle$ obsahuje první čas úmrtí v čase t_1. Druhý interval $\langle t_2, t_3 \rangle$ obsahuje čas smrti v čase t_2 a cenzorovaný čas přežití c. Poslední - třetí interval - započne v čase t_3 a obsahuje nejvyšší čas přežití, čas t_3.

Označení používaná v K – M odhadu funkce přežití

n … počet sledovaných jedinců

$t_1, t_2, \ldots, t_n$ … 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 ($r \leq n$)

$t_1 < t_2 < \ldots < t_r$ … r uspořádaných časů úmrtí

n_j … počet jedinců, kteří jsou živí před časem t_j, $j = 1, 2, \ldots, r$

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

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

Nyní předpokládejme, že smrti jedinců nastávají ve stejném okamžiku nezávisle na sobě. Kaplanův -Meierův odhad funkce přežití je dán vztahem

$$\hat{S}(t) = \prod_{j=1}^{k} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{j=1}^{k} \left(1 - \frac{d_j}{n_j} \right), \text{ kde } t_k \leq t < t_{k+1}, k = 1, 2, \ldots, r$$

Pro funkci $\hat{S}(t)$ platí $\hat{S}(t) = 1$ pro $t < t_1$.

Jak je tomu však pro $t \geq t_r$?

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, neboť nevíme, kdy přeživší jedinec zemřel za podmínky, že jeho doba přežití nebude cenzorována.

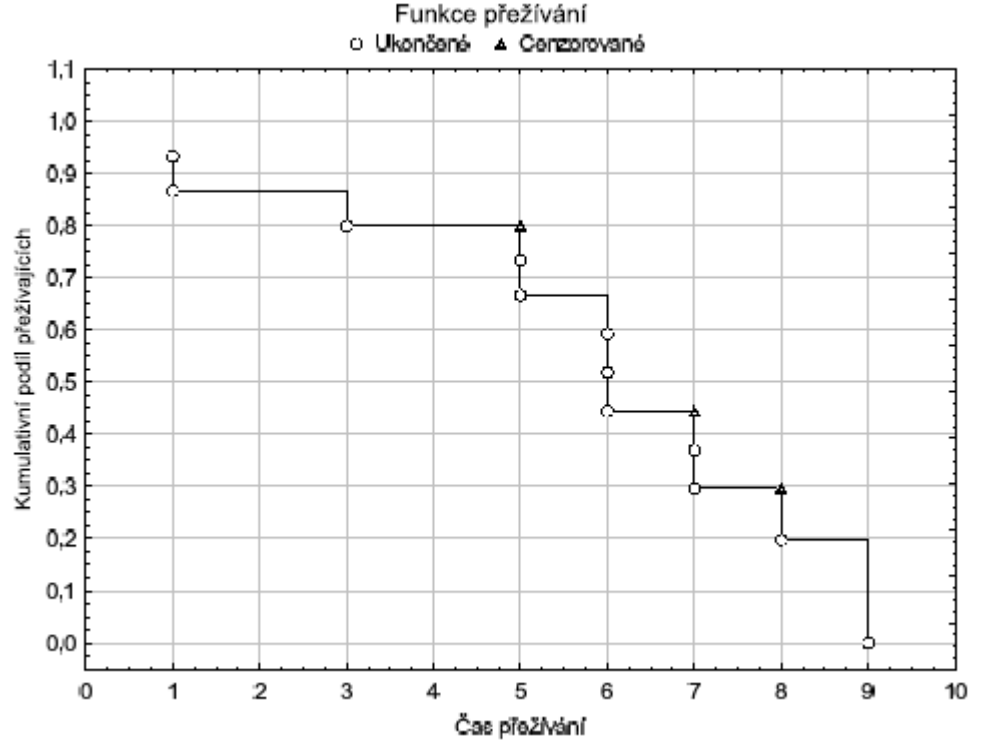
První možnost řešení: Položme $\hat{S}(t) = 0$ pro $t \geq t_r$. Tato možnost koresponduje s případem, kdy jedinec s cenzorovaným časem v t_r zemře ihned po tomto okamžiku a vede to k odhadu, který je zkreslen negativně.

Druhá možnost řešení: Položme $\hat{S}(t) = \hat{S}(t_r)$ pro $t > t_r$, což znamená, že jedinec by zemřel v čase ∞ a vede to k tomu, že odhad je zkreslen pozitivně.

Ukazuje se, že u málo početných výběrů druhá možnost je lepší. Avšak pro rozsáhlé výběry se blíží ke skutečné funkci přežití oba odhady.

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

Ukázka K – M odhadu funkce přežití



Máme 15 jedinců a 12 časů úmrtí. 3 jedinci mají cenzorované časy přežití. V čase 1 zemřeli 2 jedinci, v čase 3 jeden jedinec, v čase 5 dva jedinci a jeden je cenzorován atd.

Jednoduchá empirická funkce přežití

Nejsou-li v datovém souboru žádné cenzorované doby přežití, tj. $n_j - d_j = n_{j+1}$, $j = 1, 2, \ldots, k$, pak ze vzorce pro K – M odhad funce přežití po roznásobení dostaneme

$\hat{S}(t) = \frac{n_2}{n_1} \cdot \frac{n_3}{n_2} \cdot \ldots \cdot \frac{n_{k+}}{n_k} = \frac{n_{k+}}{n_1}$ pro $k = 1, 2, \ldots, r-1$, kde n_1 je počet jedinců, kterým hrozí smrt před prvním časem úmrtí, tedy celkový počet jedinců ve studii, a n_{k+1} je počet jedinců, jejichž čas přežití je vyšší nebo roven t_{k+1}. V důsledku toho při absenci cenzorování je odhad funkce přežití $\hat{S}(t)$ jednoduchá empirická funkce přežití definovaná jako

$\hat{S}(t)$ = počet jedinců s dobou přežití > t/ počet jedinců v souboru.

Pro funkci $\hat{S}(t)$ platí:

$\hat{S}(t) = 1$ pro $t \le t_1$, $\hat{S}(t) = 0$ pro $t > t_r$

Interval spolehlivosti pro hodnoty funkce přežití

Vyjdeme z toho, že rozptyl K – M odhadu funkce přežití je dán tzv. Greenwoodovou formulí:

$D(\hat{S}(t)) \approx \hat{S}(t)^2 \sum_{j=1}^{k} \frac{d_j}{n_j (n_j - d_j)}$ pro $t_k \le t \le t_{k+1}$.

(Greenwoodova formule podhodnocuje reálný rozptyl K- M odhadu při malých rozsazích výběru.)

100(1-α)% asymptotický interval spolehlivosti pro hodnotu S(t) funkce př

$d = \hat{S}(t) - \hat{S}(t) \sqrt{\sum_{j=} \frac{d_j}{n_j (n_j - d_j)}} \cdot u_{1-\alpha/2}$

$h = \hat{S}(t) + \hat{S}(t) \sqrt{\sum_{j=} \frac{d_j}{n_j (n_j - d_j)}} \cdot u_{1-\alpha/2}$

$\hat{S}(t)$

$\hat{S}(t)$

Testování hypotézy o rozdílu mezi dvěma a více skupinami

V analýze přežití se samozřejmě 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. Zde se seznámíme s log-rank testem, Gehanovým – Wilcoxonovým testem a chí-kvadrát testem.

Případ dvou skupin

Máme dvě skupiny jedinců. Předpokládáme, že v obou skupinách dohromady je r časů úmrtí, $t_1 < t_2 < \ldots < t_r$ 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.

Oba zde popsané testy - log-rank test i 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, …, r.

Log-rank test

Zavedeme statistiku $U_L = \sum_{j=1}^{r} (d_{1j} - e_{1j})$. Její rozptyl $D(U_L)$ je dán vztahem $D(U_L) = \sum_{j=1}^{r} \frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j^2 (n_j - 1)}$. 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 α, když $W_L \geq \chi^2_{1-\alpha}(1)$.

Gehanův – Wilcoxonův test

Zavedeme statistiku $U_W = \sum_{j=1}^{r} n_j (d_{1j} - e_{1j})$. (Vidíme, že rozdíl mezi U_L a U_W spočívá v tom, že ve statistice U_W je každý rozdíl $d_{1j} - e_{1j}$ vynásoben vahou n_j, což je počet jedinců v obou skupinách, kteří mohou zemřít v čase t_j, $j = 1, 2, \ldots, r$. V důsledku toho je rozdílu $d_{1j} - e_{1j}$ kladena menší váha v čase t_j, pokud počet žijících je malý, tedy například při j blížící se r. To znamená, že statistika U_W je méně citlivá na odchylky d_{1j} od e_{1j} při nejpozdějších časech úmrtí.) Její rozptyl $D(U_W)$ je dán vztahem $D(U_W) = \sum_{j=1}^{r} \frac{n_{1j} n_{2j} d_j (n_j - d_j)}{n_j - 1}$. 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 α, když $W_W \geq \chi^2_{1-\alpha}(1)$.

Srovnání Wilcoxonova a log-rank testu

Při testování nulové hypotézy, že neexistuje rozdíl mezi funkcemi přežití dvou skupin jedinců, je dobré vědět, který z uvedených dvou testů je vhodnější použít.
Jednoduchou pomůckou, jak to zjistit, je vykreslit si obě funkce přežití do jednoho grafu a podle jejich průběhu vybrat vhodnější test. Pokud se například tyto dvě funkce přežití kříží, je vhodnější k testování nulové hypotézy zvolit Gehanův - Wilcoxonův test. Pokud je tomu však naopak a funkce se nekříží, je lépe zvolit log-rank test.

Případ tří a více skupin

Nyní rozšíříme výsledky, ke kterým jsme dospěli pro dvě skupiny. 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.

Zavedeme statistiky $U_{Lk} = \sum_{j=1}^{r} \left\{ d_{kj} - \frac{n_{kj} d_j}{n_j} \right\}$, $U_{Wk} = \sum_{j=1}^{r} n_j \left\{ d_{kj} - \frac{n_{kj} d_j}{n_j} \right\}$, k = 1, 2, …, p-1. Tyto statistiky uspořádáme do sloupcových vektorů U_L a U_W, každý má p-1 složek. Kovariance mezi statistikami U_{Lk} a $U_{Lk'}$ je dána vztahem

$v_{Lkk'} = \sum_{j=1}^{r} \frac{n_{kj} d_j (n_j - d_j)}{n_j (n_j - 1)} \left\{ \delta_{kk'} - \frac{n_{k'j}}{n_j} \right\}$ pro k, k‘ = 1, 2, …, p-1, kde $\delta_{kk'} = \begin{cases} 1 \text{ pro } k = k' \\ 0 \text{ jinak} \end{cases}$. Tyto kovariance můžeme zapsat do varianční matice V_L řádu p-1, což je symetrická matice, která má na hlavní diagonále rozptyly a mimo diagonálu má kovariance. Tedy

$$V_L = \begin{pmatrix} V_{L11} & V_{L12} & \cdots & V_{L1(p-1)} \\ V_{L21} & V_{L22} & \cdots & V_{L2(p-1)} \\ \vdots & \vdots & \ddots & \vdots \\ V_{L(p-1)1} & V_{L(p-1)2} & \cdots & V_{L(p-1)(p-1)} \end{pmatrix},$$

kde V_{Lij} je rozptyl $D(U_{Li})$ pro i = j a kovarince $C(U_{Li}, U_{Lj})$ pro $i \neq j$, i, j = 1, 2, …, p-1. Podobně (k,k’)-tý prvek varianční matice V_W pro Gehanovu – Wilcoxonovu statistiku je dán vztahem $v_{Wkk'} = \sum_{j=1}^{r} n_j^2 \frac{n_{kj} d_j (n_j - d_j)}{n_j (n_j - 1)} \left\{ \delta_{kk'} - \frac{n_{k'j}}{n_j} \right\}$.

Ve výsledku získáme testovou statistiku $U_L' V_L^{-1} U_L$ nebo $U_W' V_W^{-1} U_W$. V případě platnosti nulové hypotézy se testová statistika asymptoticky řídí rozložením $\chi^2(p-1)$. Nulovou hypotézu tedy zamítáme na asymptotické hladině významnosti α, když $U_L' V_L^{-1} U_L \geq \chi^2_{1-\alpha}(p-1)$ resp. $U_W' V_W^{-1} U_W \geq \chi^2_{1-\alpha}(p-1)$.

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.

Ukázka části datového souboru:

Data: slinivka.sta* (6s krát 67ř)

	1 věk	2 sex	3 přežití	4 smrt	5 stadium	6 typ
1	68	muz	1	nezije	3	duktální adenokarcinom
2	65	zena	1	nezije	3	ampulární karcinom
3	75	muz	3	nezije	3	ampulární karcinom
4	43	zena	5	nezije	3	duktální adenokarcinom
5	63	zena	5	nezije	3	karcinom terminálního choledochu
6	55	zena	5	nezije	4	karcinom terminálního choledochu
7	73	zena	6	nezije	3	duktální adenokarcinom
8	65	zena	6	nezije	3	duktální adenokarcinom
9	59	muz	6	nezije	3	duktální adenokarcinom
10	49	zena	7	nezije	4	duktální adenokarcinom
11	45	muz	7	zije	3	duktální adenokarcinom
12	74	zena	7	nezije	3	ampulární karcinom
13	72	zena	7	nezije	2	karcinom terminálního choledochu
14	67	muz	8	nezije	2	duktální adenokarcinom
15	59	zena	8	nezije	3	karcinom terminálního choledochu
16	51	muz	9	nezije	2	duktální adenokarcinom
17	60	muz	9	nezije	2	duktální adenokarcinom
18	51	muz	9	nezije	4	duktální adenokarcinom
19	65	muz	9	nezije	3	karcinom terminálního choledochu
20	71	zena	9	zije	2	karcinom terminálního choledochu
21	57	zena	10	nezije	3	duktální adenokarcinom
22	53	zena	10	nezije	3	duktální adenokarcinom
23	54	muz	10	nezije	3	karcinom terminálního choledochu

Popis proměnných:

Věk … věk pacienta v letech

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

Přežití … doba přežití v měsících

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

Stadium … stadium rakoviny, 1 až 4

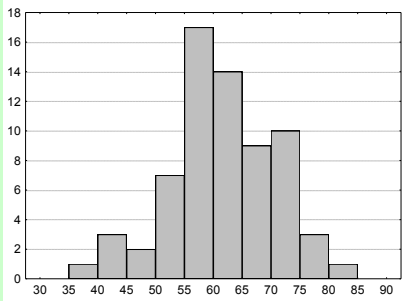
Typ … typ nádoru (1 - duktální adenokarcinom, 2 - ampulární karcinom, 3 - karcinom terminálního choledochu)

Výpočet pomocí systému STATISTICA

Číselné charakteristiky věku

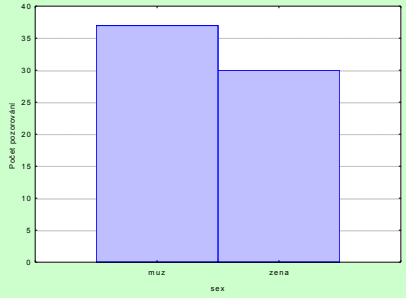
Proměnná	N platných	Průměr	Medián	Minimum	Maximum	Sm. odch.	Šikmost	Špičatost
věk	67	62,1	63	36	81	9,18	-0,36	0,20

Histogram věku



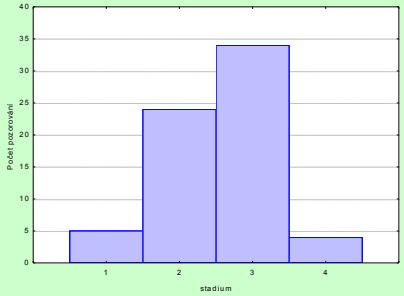
Četnostní tabulka proměnné sex

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



Četnostní tabulka proměnné stadium

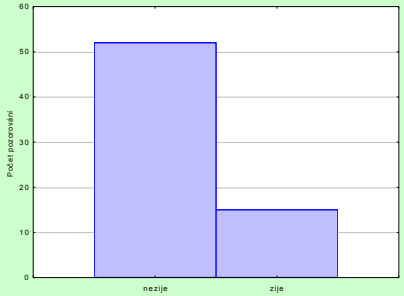
Kategorie	Tabulka četností: stadium (sliniv ka.sta)			
	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
1	5	5	7,46	7,46
2	24	29	35,82	43,28
3	34	63	50,75	94,03
4	4	67	5,97	100,00



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

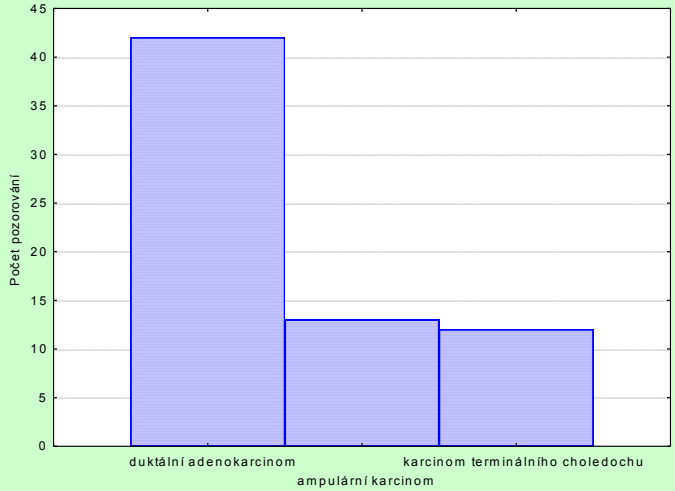
Kategorie	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
zije	15	15	22,39	22,39
nezije	52	67	77,61	100,00



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
tě

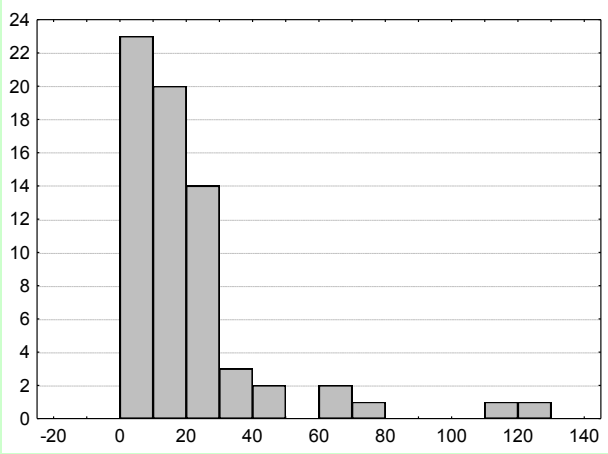
Číselné charakteristiky proměnné přežití (v měsících)

Proměnná	Průměr	Minimum	Maximum	Sm. odch.	Šikmost	Špičatost
přežití	21,87	1	121	22,84	2,85	9,45

Číselné charakteristiky proměnné přežití (v měsících) – pokračování

Proměnná	Medián	Spodní kvartil	Horní kvartil	Kvartilové rozpětí
přežití	14	9	27	18

Histogram proměnné přežití



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 přežití, 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.

	Popisné statistiky pro každou skupinu (slinivka.sta)					
Skupina	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 karcinomem 36,8 měsíců a jedinci, u kterých se vyskytl karcinom terminálního choledochu, se v průměru dožívají 26 měsíců.

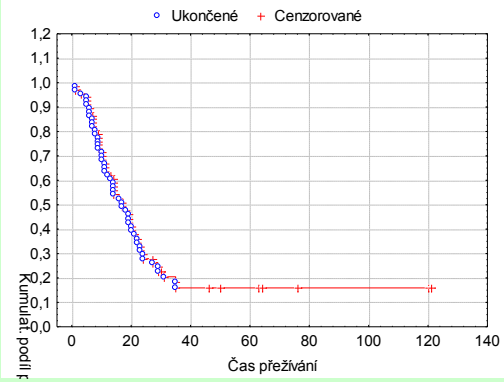
Výsledky pro proměnnou sex

	Popisné statistiky pro každou skupinu (slinivka.sta)					
Skupina	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

Výsledky pro proměnnou smrt

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

Nyní získáme Kaplanův – Meierův odhad funkce přežití pro celý soubor:
Statistiky – Pokročilé lineární/nelineární modely – Analýza přežívání – Kaplan – Meierova metoda - Proměnné – Přežívání přežití, Cenzor. prom. 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 přežití, Cenzor. prom. smrt – OK – Kódy pro ukončené nežije, Kódy pro cenzorované žije, Kódy skupin Vše – OK – OK – záložka Detaily – Kvantily funkce přežívání

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

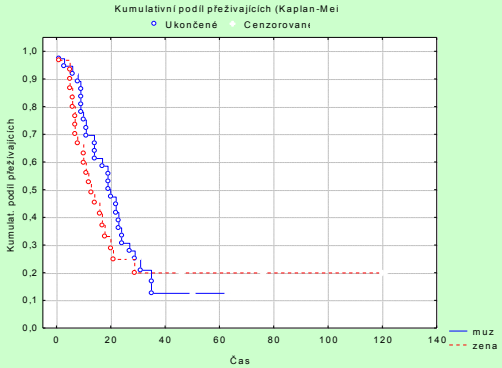
Z těchto údajů vyplývá, že jedna čtvrtina pacientů umírá do 9. měsíce od zjištění nádoru, polovina jedinců zemře do 17. měsíce, a přibližně do 28,4 měsíců zemřou tři čtvrtiny jedinců.

Porovnání doby přežití 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) Četnost označených buněk > 10 (Marginální součty nejsou označeny)			
	sex	smrt zije	smrt nezije	Řádk. součty
Četnost	muz	7	30	37
Řádk. četn.		18,92%	81,08%	
Četnost	zena	8	22	30
Řádk. četn.		26,67%	73,33%	
Četnost	Vš.skup.	15	52	67

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 přežití, Cenzor. prom. smrt, Grupovací prom. sex – OK – Kódy pro ukončené nežije, Kódy pro cenzorované žije – OK – záložka Grafy funkcí – Kum. podíl přežív. 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 $\alpha = 0,$

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

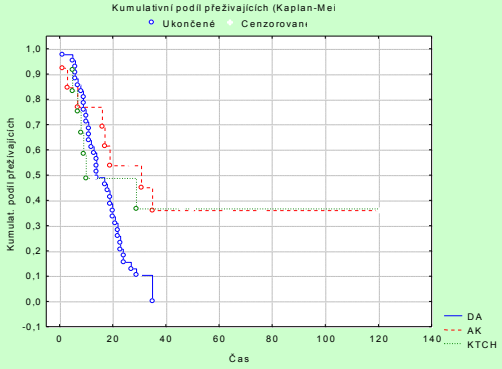
Hodnota testové statistiky je rovna 1,5535 a příslušná p-hodnota je 0,1203. Jelikož je p-hodnota větší než 0,05, tak hypotézu H_0 nezamítáme na asymptotické hladinì významnosti 0,05. Tudíž

Srovnání doby přežití 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 typ a smrt.

	Kontingenční tabulka (slinivka.sta) Četnost označených buněk > 10 (Marginální součty nejsou označeny)			
	typ	smrt zije	smrt nezije	Řádk. součty
Četnost	duktální adenokarcinom	5	37	42
Řádk. četn.		11,90%	88,10%	
Četnost	ampulární karcinom	5	8	13
Řádk. četn.		38,46%	61,54%	
Četnost	karcinom terminálního choledochu	5	7	12
Řádk. četn.		41,67%	58,33%	
Četnost	Vš.skup.	15	52	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 přežití, Cenzor. prom. smrt, Grupovací prom. typ – OK – Kódy pro ukončené nežije, Kódy pro cenzorované žije – OK – záložka Grafy funkcí – Kumul. podíl přežív. (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

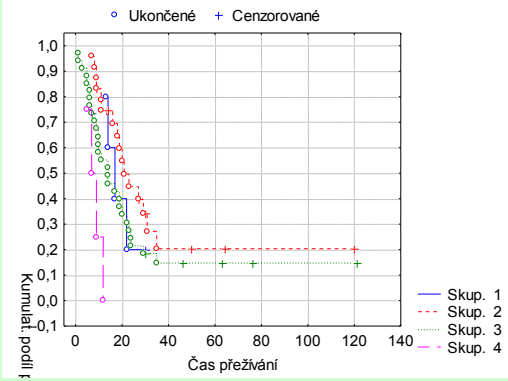
Hodnota testové statistiky je 2,0437 a příslušná p-hodnota je 0,3599. Poněvadž je p-hodnota větší než 0,05, tak nulovou hypotézu o shodě dob přežití u jednotlivých typů karcinomů nezamítáme na asymptotické hladině významnosti 0,05.

Srovnání doby přežití 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 stadium a smrt.

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