

Analýza přežívání čertic a čertů

Ing. Milan Němeček

Vzpomeňme si na pohádku s Čerty nejsou žerty. V ní Lucifer (dále jen Lúca) pověřil čerta Janka, aby přinesl Dorotu Máchalovou do pekla, poněvadž míra jejích hříchů přesáhla přípustnou hranici. Předtím než Janek vyrazil na cestu, dostal od Lúci zevrubný plán vesnice a následující varování: „Pozor na peří a žádný alkohol.“ První varování bylo v příběhu náležitě vysvětleno a my nyní zkusíme odhadnout, proč to druhé.

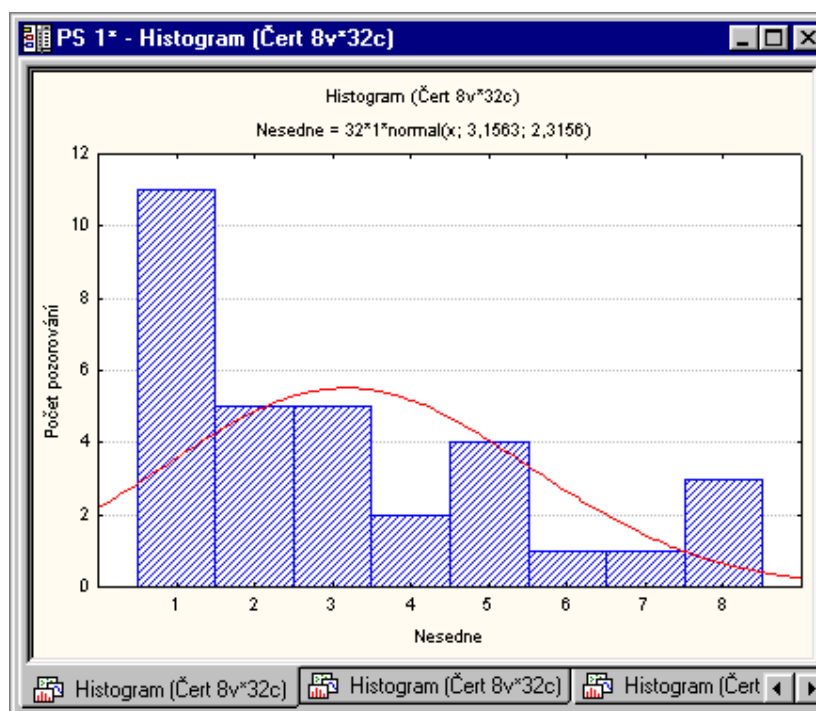
Večer pátého prosince Mikuláš s andělem a čertem nebo nádhernou čerticí obcházejí domácnosti a rozdávají dětem dárky. Přiznejme si, čerti a čertice působí jako výchovný prvek. Na nejmenší ratolesti určitě, u náctiletých působí už poněkud jinak. Za přinesené dárky a „výchovné působení“ často dostane návštěva z nebes i z pekla něco pro zahřátí na jejich další cestu. Stává se, že chudáci z pekla se někdy musí za někoho ze skupiny „obětovat“, například za anděla, a mohou jim nastat problémy při plnění jejich povinností. Byl proveden reálný průzkum se základním cílem odhadnout bezporuchovost čerta v závislosti na počtu přijatých odměn. Pro následnou analýzu bylo shromážděno třicet dva záznamů, viz tabulka níže, jedná se tedy o reálná data.

Data: Čert* (8s krát 32ř)								
Čerti a čertice								
	1	2	3	4	5	6	7	8
	Pohlaví	Nesedne	Normál	Extrém	Stát	Oblast	Věk	Cenzorováno
1	Čert	4	7	10	CR	Olomouc	30	ne
2	Čert	2	6	8	SR	Nove Zamky	30	ne
3	Čert	3	9	18	CR	České Budějovice	20	ne
4	Čert	2	6	20	CR	České Budějovice	30	ne
5	Čert	5	7	10	CR	Havlíčkův Brod	50	ne
6	Čert	1	3	4	CR	Praha	40	ne
7	Čert	1	2	4	CR	Praha	60	ne
8	Čert	8	10	16	CR	Plzeň	30	ne
9	Čert	1	1	1	CR	Svitava	20	ne
10	Čert	1	1	1	CR	Praha	40	ne
11	Čert	6	12	20	CR	Praha	30	ne
12	Čert	8	12	25	CR	Praha	30	ne
13	Čert	8	10	20	SR	Poprad	50	ne
14	Čert	5	7	10	CR	Praha	30	ne
15	Čert	5	10	20	CR	Kolín	30	ne
16	Čert	5	10	20	CR	Praha	30	ne
17	Čert	3	6	14	CR	Praha	20	ne
18	Čert	3	5	9	CR	Nymburk	50	ne
19	Čertice	1	3	7	SR	Topolčany	60	ne
20	Čertice	1	6	15	CR	Praha	40	ne
21	Čertice	2	4	8	CR	Plzeň	20	ne
22	Čertice	1	8	10	CR	Plzeň	30	ne
23	Čertice	2	4	7	CR	Praha	40	ne
24	Čertice	1	6	11	CR	Hradec Králové	20	ne
25	Čertice	1	3	5	SR	Piešťany	30	ne
26	Čertice	3	6	8	SR	Trenčín	20	ne
27	Čertice	4	8	10	CR	Praha	20	ne
28	Čertice	1	3	4	CR	Praha	30	ne
29	Čertice	1	2	5	CR	Praha	50	ne
30	Čertice	7	9	14	CR	Kolín	30	ne
31	Čertice	2	5	8	CR	Praha	20	ne
32	Čertice	3	7	15	SR	Lipt. Mikuláš	40	ano

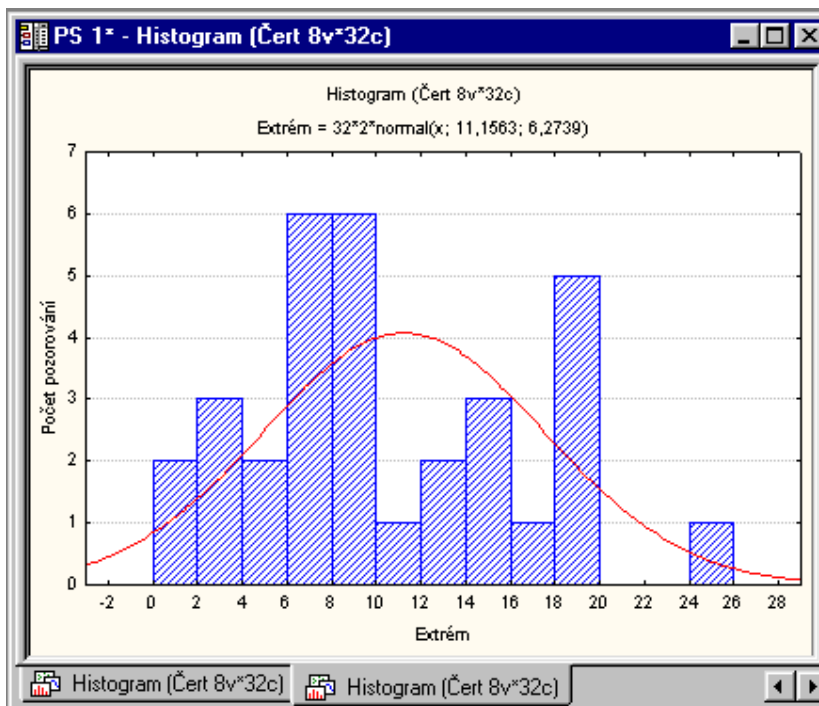
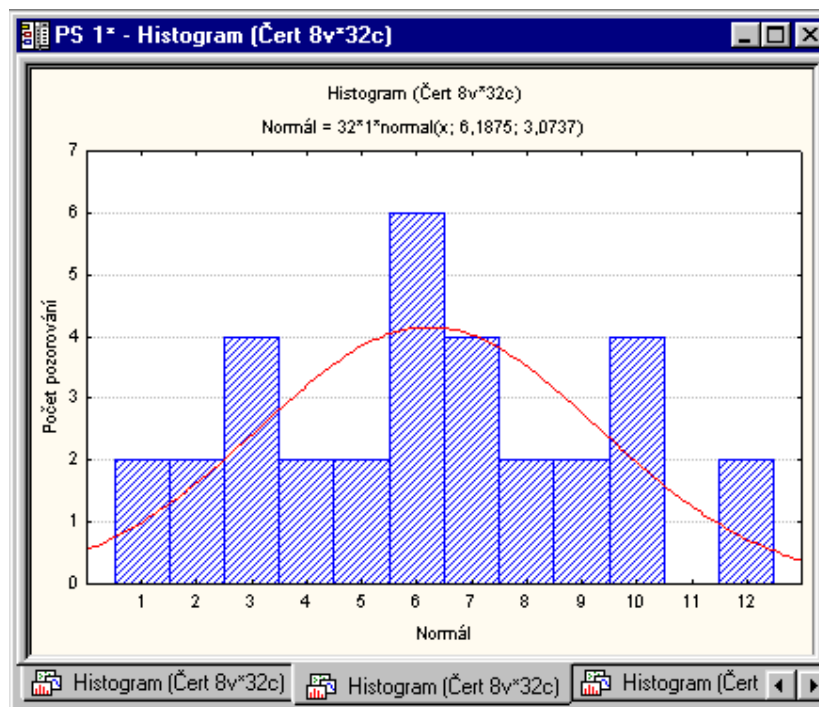
Každá bytost byla požádána, aby uvedla nebo odhadla, kolik panáků je schopná vypít, když není v kondici (zkrátka jí to nesedne), potom byla požádána o údaj při jejím normálním stavu. Na závěr odhadla (nebo uvedla ze zkušenosti) extrémní množství odměn, které je schopna přežít a nepadnout za peklo. Samozřejmě, někteří jedinci nechtěli uvést údaje, ale přesto se povedlo shromáždit malý soubor dat, v němž je možné zkusit nalézt určité informace. Pokud vás to bude zajímat, čertice byly sdílnější.

Výše uvedené údaje obsahují určitou míru nejistoty, a proto je nutné použít metody statistické analýzy. Zvolíme neparametrické metody. Těm dáváme přednost při malém počtu dat, poněvadž není možné aplikovat centrální limitní větu na výběrový průměr. Dále můžeme předpokládat výrazně nenormální rozdělení. Zároveň se nám v záznamech vyskytuje cenzorovaný údaj. Program *STATISTICA* ve svém modulu *Pokročilé lineární/nelineární modely* nabízí skupinu metod spadající do oblasti *Analýzy přežívání*. Tyto metody začaly být poprvé používány v medicíně a biologii, velmi brzo však našly uplatnění i v technických oborech, například v oblasti spolehlivosti.

Počet záznamů je velmi malý, v průzkumové analýze dat bude dostačující provést pouze jejich grafické zobrazení se zakreslenou křivkou aproximujícího normálního rozdělení. Pokud daná bytost (tedy neuvažujeme pohlaví) není v kondici, takto vypadá rozložení četnosti hodnot v souboru. Na vodorovné ose je počet panáků, při kterých je bytost ještě schopna si plnit své povinnosti.

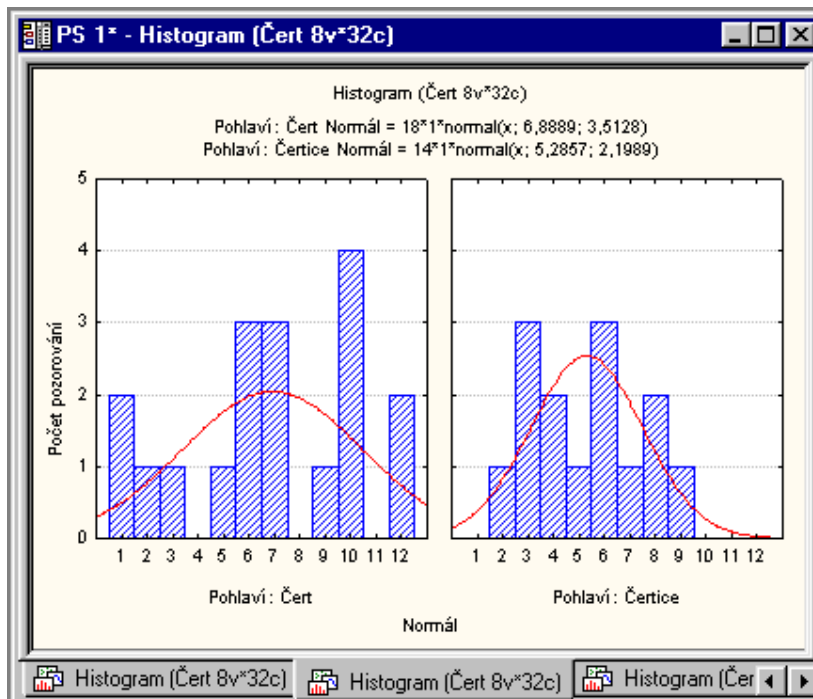
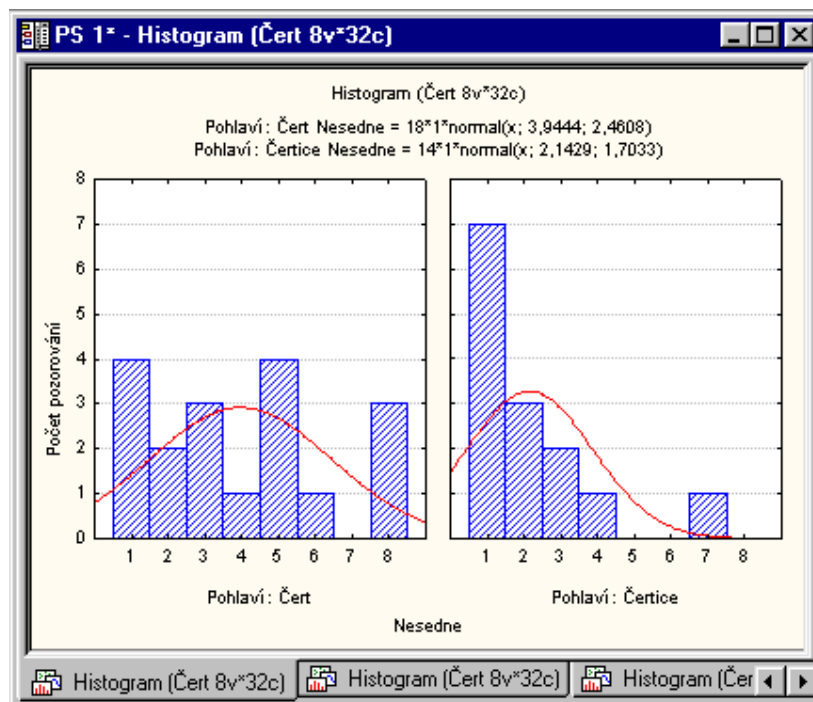


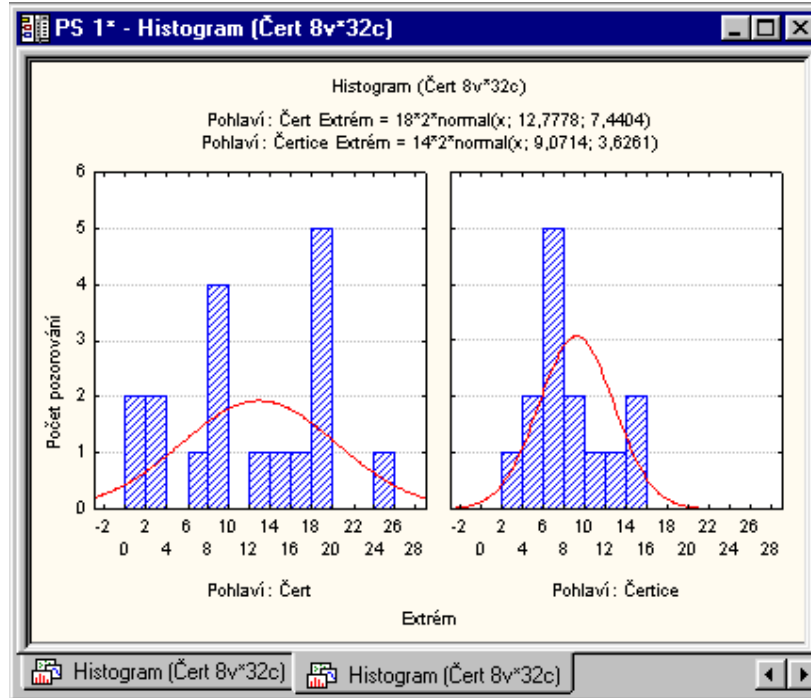
Jak vidíme, rozložení hodnot se blíží k exponenciálnímu rozdělení. Níže jsou uvedeny zbývající grafy.



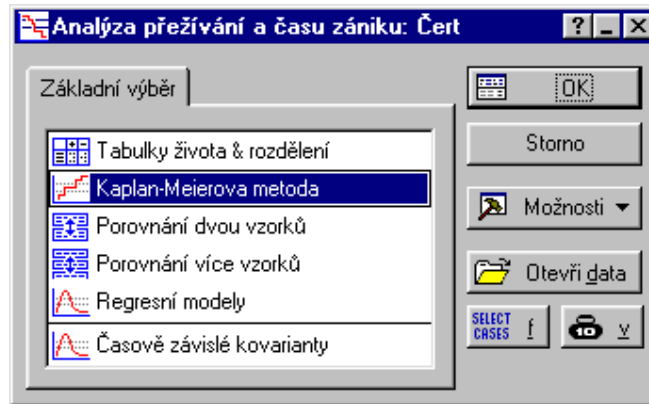
U posledního grafu vidíme součet dvou různých rozdělení. Jejich oddělení by bylo možné pouze na základě nějakého atributu, např. oblast, pohlaví apod. Stále však musíme mít na paměti, že pracujeme s velmi malým počtem dat, tento předpoklad však při větším rozsahu záznamů nemusí platit nebo může být potvrzen.

Samozřejmě jsme už zvědaví na rozdíl mezi „mužem“ a „ženou“.





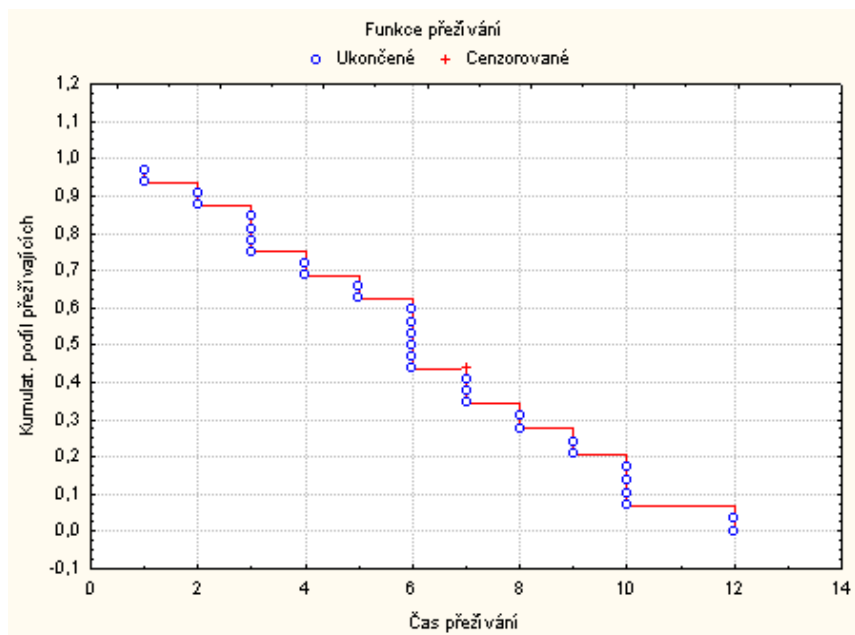
Například v krabicovém grafu bychom lépe viděli odlehle hodnoty nebo extrémy, ale vraťme se nyní k neparametrickým metodám. Program *STATISTICA* nabízí níže zobrazené metody z oblasti *Analýzy přežívání*.



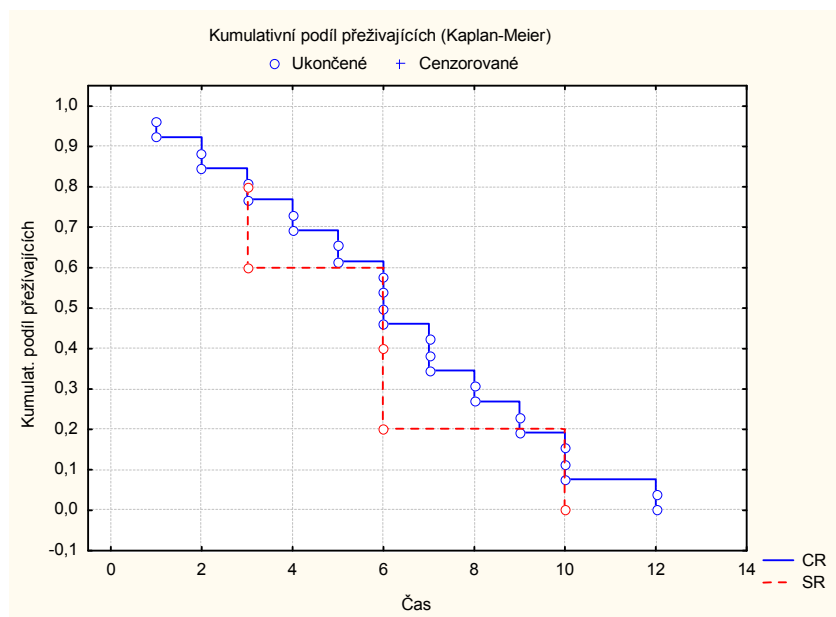
Kaplan-Meierova metoda patří do skupiny neparametrických metod. Je upřednostňována před výpočty pomocí tabulky života vzhledem k jejím přesnějším výstupům. Obdobou Kaplan-Meierovy metody je Nelson-Altschulerův odhad, jenž však dává lepší výsledky, než je skutečnost. Bez rozlišení pohlaví, státní, věku a za normální kondice čertovské bytosti získáme níže uvedené kvantily doby přežití.

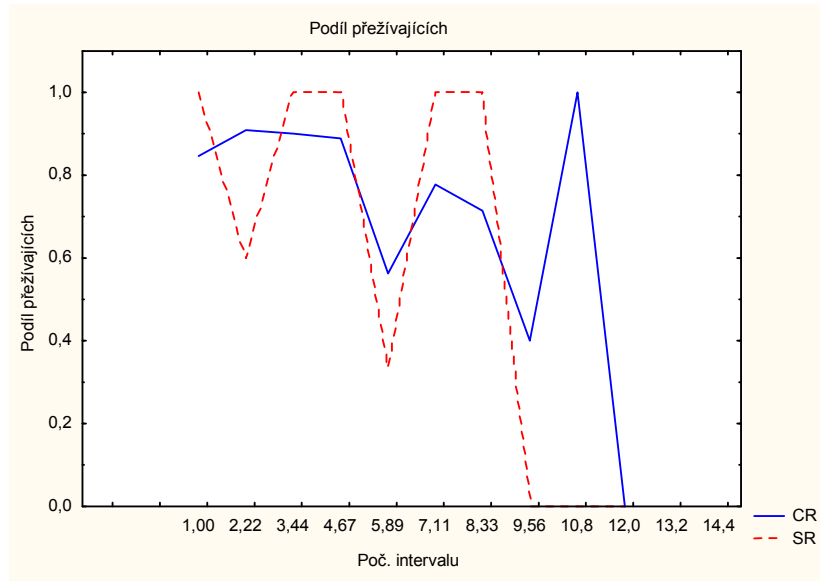
Kvantily	Kvantily (Čert) funkce přežívání	
		Čas přeživ.
25. kvantil (dolní kvartil)		3,000000
50. kvantil (medián)		6,000000
75. kvantil (horní kvartil)		8,250000

Dá se předpokládat, že padesát procent „pracujících“ vydrží 6 naturálií v době mikulášských pochůzek. Vypočtené hodnoty jsou přehledněji zobrazeny v Kaplan-Meierově grafu.



Zkusme zjistit, zda existuje statisticky významný rozdíl mezi „zajištěním služeb“ v České republice a na Slovensku.

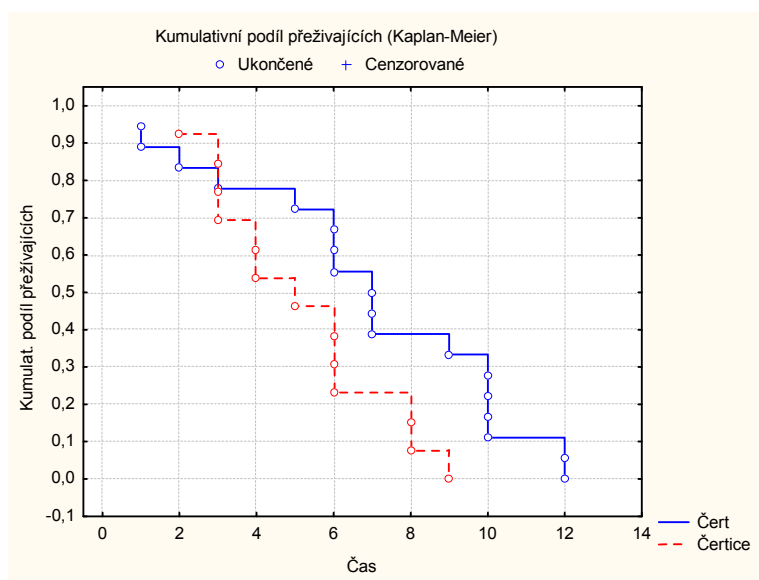
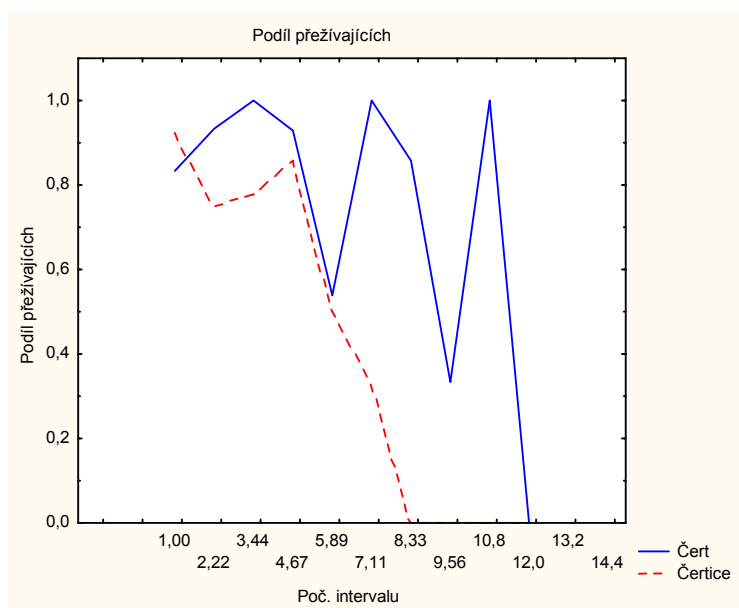




Při malém počtu dat a za předpokladu jejich proložení exponenciálním nebo Weibullovým rozdělením se doporučuje používat Coxův F-test, který má v těchto případech podstatně větší sílu testu jak obvykle používaný Gehanův-Wilcoxonův test. Za nulovou hypotézu je považováno tvrzení, které deklaruje „žádný rozdíl“. Tedy - jakýkoliv nalezený rozdíl mezi jednotlivými skupinami je způsoben přirozenou variabilitou dat. Hladina významnosti alfa je pravděpodobnost zamítnutí nulové hypotézy při její platnosti, standardně se volí hodnota 0,05.

Přesné ukončení	Coxův F-test (Čert)			
	R(l)	M(l)	M/R	Kap/Meir odhad
1,0000	31,00000	2,000000	0,064516	1,000000
2,0000	29,00000	2,000000	0,068966	0,935484
3,0000	27,00000	4,000000	0,148148	0,870968
4,0000	23,00000	2,000000	0,086957	0,741935
5,0000	21,00000	2,000000	0,095238	0,677419
6,0000	19,00000	6,000000	0,315790	0,612903
7,0000	13,00000	3,000000	0,230769	0,419355
8,0000	10,00000	2,000000	0,200000	0,322581
9,0000	8,00000	2,000000	0,250000	0,258064
10,000	6,00000	4,000000	0,666667	0,193548
12,000	2,00000	2,000000	1,000000	0,064516
				0,000000

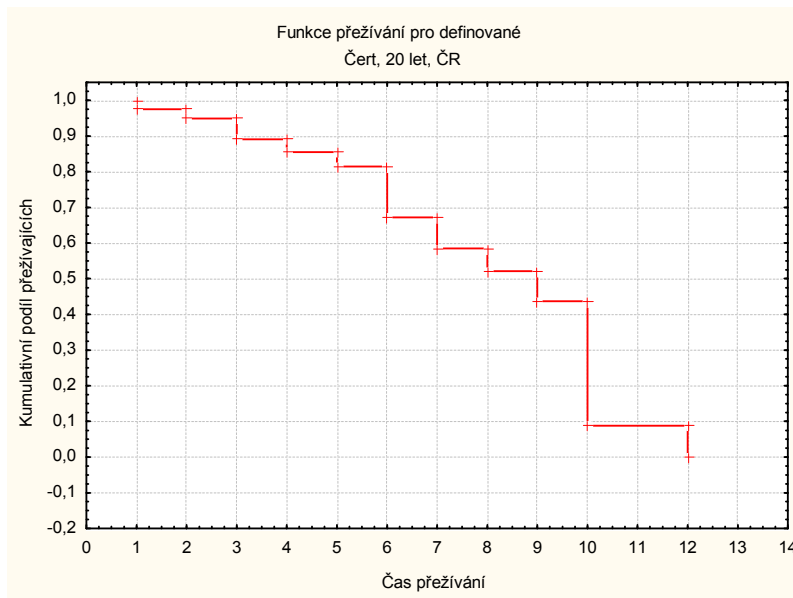
Podle vypočtené p-hodnoty $> 0,05$ je možné přijetí nulové hypotézy. Mezi oběmi skupinami vzorků dat není statisticky významný rozdíl. Porovnejme čerta a čertici v normální kondici bez přihlídnutí k jejich státní působnosti.



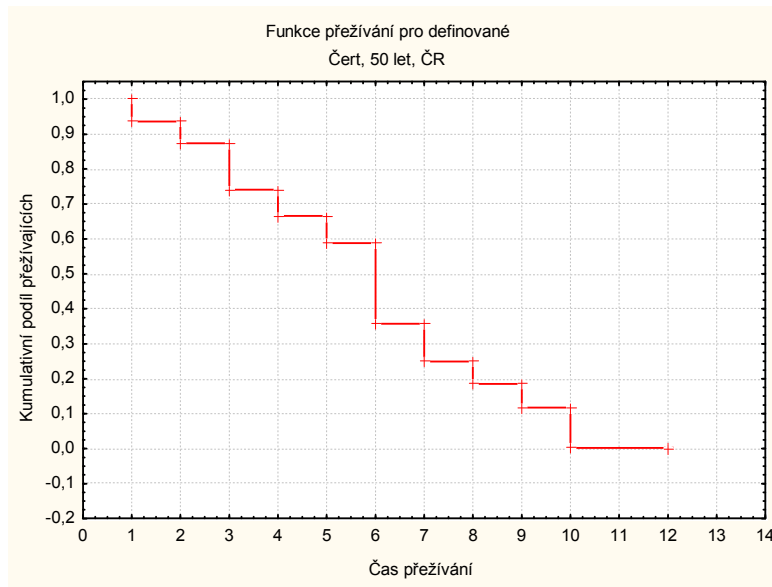
Přesné ukončení	Coxův F-test (Čert) T1 = 22,95042 T2 = 8,049580 F(36, 26) = 2,059151 p = ,02920			
	R(l)	M(l)	M/R	Kap/Meir odhad
1,0000	31,00000	2,000000	0,064516	1,000000
2,0000	29,00000	2,000000	0,068966	0,935484
3,0000	27,00000	4,000000	0,148148	0,870968
4,0000	23,00000	2,000000	0,086957	0,741935
5,0000	21,00000	2,000000	0,095238	0,677419
6,0000	19,00000	6,000000	0,315790	0,612903
7,0000	13,00000	3,000000	0,230769	0,419355
8,0000	10,00000	2,000000	0,200000	0,322581
9,0000	8,00000	2,000000	0,250000	0,258064
10,000	6,00000	4,000000	0,666667	0,193548
12,000	2,00000	2,000000	1,000000	0,064516
				0,000000

Jak jsme tušili, potvrdil se statisticky významný rozdíl ve výdrži čertů a čertic.

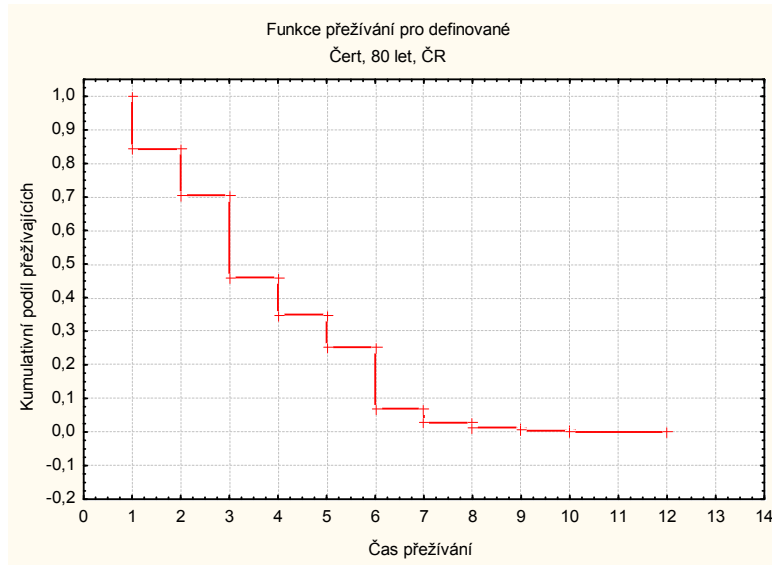
Velmi zajímavou informací bude posouzení vlivu věku na výdrž pro blíže specifikovanou bytost. Pro analýzu dat je použita metoda Coxovy regrese s proporcionálním rizikem, tedy s konstantními kovariantami. Opět k analýze jsou použity údaje o výdrži bytostí v normální kondici. Jako nezávislé proměnné (kovarianty) zvolíme věk, stát a pohlaví. V programu *STATISTICA* byl sestaven příslušný model a podle vložených hodnot je proveden grafický výstup.



Takže slečny a paní, na základě analyzovaných dat můžete předpokládat, že 90 procent čertů bude pravděpodobně ještě v kondici při třech panáčích.

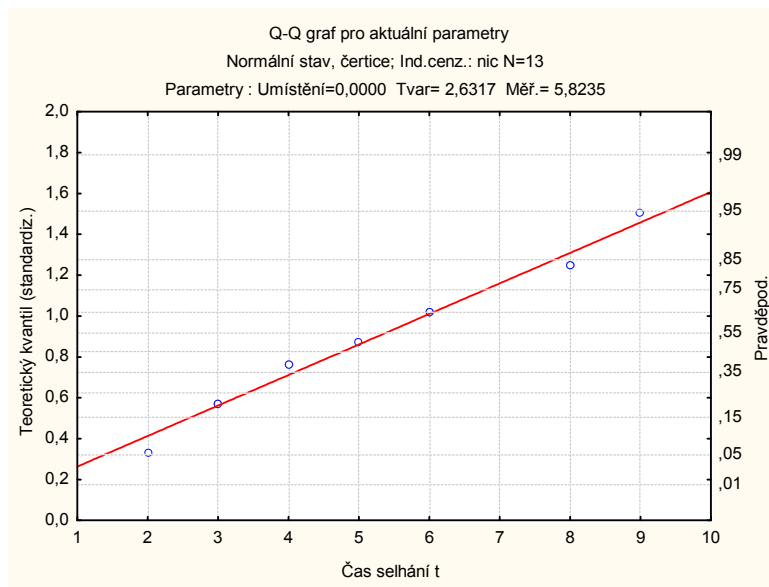
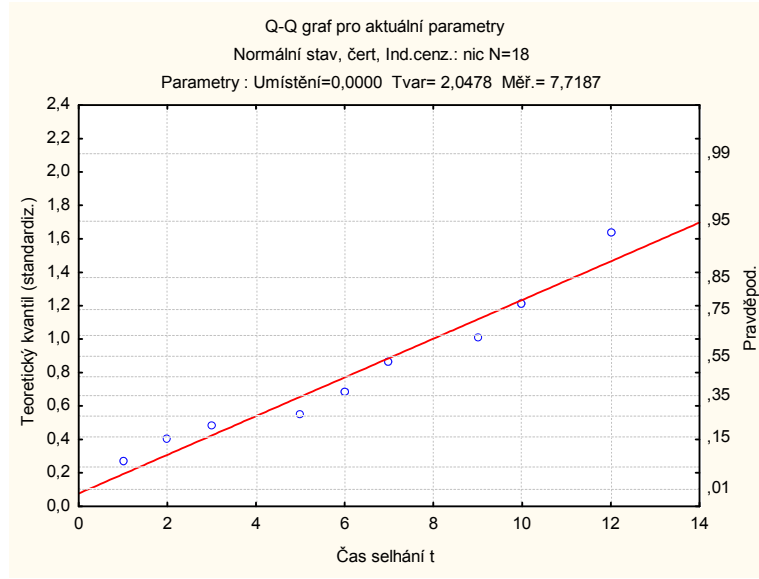


A padesát procent padesátiletých čertů po pěti až šesti panáčích bude na své horní hranici nebo už bude mít „poruchu“.



A se starším čertem budete muset trochu opatrněji nebo se z vás stane první pomoc. Co vy na to? Všimněme si, že v datech se nevyskytuje údaj o „kvalifikaci“ osmdesátiletého čerta. Na základě sestaveného modelu můžeme provést alespoň odhad.

Při porovnávání dvou vzorků dat byl použit Coxův F-test s předpokladem aproximace dat Weibullovým rozdělením. Potvrďme daný předpoklad například pro kategorii čertů a čertic bez rozlišení věku, státní příslušnosti a za jejich standardní kondice. Program *STATISTICA* ve své nabídce průmyslových statistik nabízí metodu jak grafického, tak i numerického řešení odhadu parametrů Weibullova rozdělení a souvisejících statistik a výpočtů. Pro jednotlivé kategorie dat je provedeno ověření za pomoci Q-Q grafu.

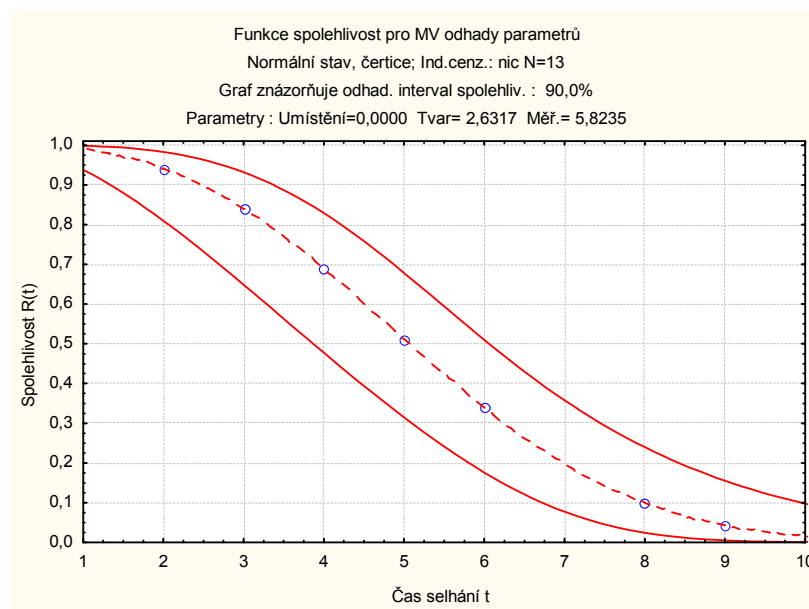
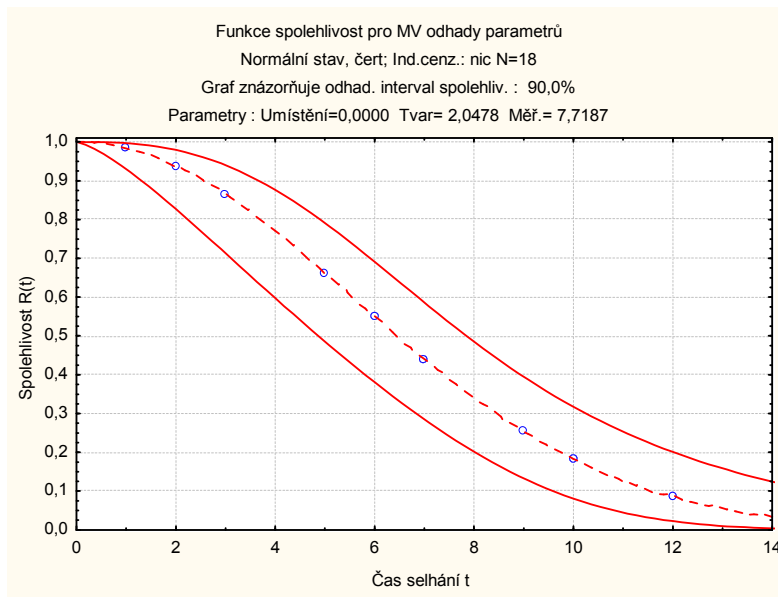


Data jsou proložena přímkou, jejíž směrnice odpovídá tvaru dvou-parametrového rozdělení a měřítko je dáno průsečíkem této přímky s hodnotou 63,2 procent pravděpodobnosti. Z rozmístění bodů kolem přímky můžeme usuzovat na vhodnost aproximace dat Weibullovým rozdělením. Vypočtené parametry rozdělení jsou uvedeny v podnadpisu grafu.

Test	Test	p
	Hodnota	
Hollander-Proschan	-0,355860	p=,72195
Mann-Scheuer-Fertig	0,351963	p>,25
Anderson-Darling	0,754094	p<,05

Testy kvality proložení (Čert)		
Normální stav; Ind.cenz.: nic N=13		
Parametry : Umístění=0,0000 Tvar= 2,6317 Měř.= 5,8235		
Filtr pro zahrnutí: v1='Čertice'		
Test	Test Hodnota	p
Hollander-Proschan	0,104543	p=,91674
Mann-Scheuer-Fertig	0,515814	p>.25
Anderson-Darling	0,366368	p>.20

Numerické testy rovněž potvrzují shodu dat s předpokládaným Weibullovým rozdělením. Pro jednotlivé kategorie je zobrazen graf spolehlivosti čerta nebo čertice na době selhání (v našem případě počtu panáků) s vyznačeným 90ti procentním intervalem spolehlivosti.



Co říci na závěr? Shrňme si základní informace. Analyzovaná data bylo možné aproximovat Weibullovým rozdělením nebo modelem Coxovy regresní analýzy s proporcionálním rizikem. Není statistický významný rozdíl mezi „zajištěním služeb“ v České republice a na Slovensku. Obecně mají čerti větší výdrž jak čertice. Pokud jsou v normální kondici, dá se předpokládat u čertic 50ti procentní výdrž při pěti zkonsumovaných naturáliích, u čertů přibližně 6 až 7. Budete-li se o Vánocích dívat na pohádku zmíněnou v úvodu článku, odhadněte věk a kondici čerta Janka. Použijte tyto údaje jako vstupní parametry zmíněné regresní analýzy a odhadněte, kolik kaprála stál otisk palce čerta Janka při odvodovém řízení v hospodě. O trial verzi programu *STATISTICA* můžete požádat u firmy StatSoft. Všem čerticím a čertům děkuji za poskytnutá data, jim a čtenářům tohoto článku přeji hezké prožití vánočních svátků, divokého Silvestra a do dalšího roku hodně zdraví, pohody a splněných přání.