7 VYBRANÁ ROZDĚLENÍ SPOJITÉ NÁHODNÉ VELIČINY

Pro nezápornou náhodnou veličinu X se spojitým rozdělením definujeme pro $F(t)\neq 1$ (tj. $F(t)<1$) **intenzitu poruch $\lambda(t)$** :

$$\lambda(t)=\frac{f(t)}{1-F(t)}$$

Představuje-li náhodná veličina X **dobu do poruchy** nějakého zařízení, pak intenzita poruch vyjadřuje, že pokud do času t nedošlo k žádné poruše, tak pravděpodobnost, že k ní dojde v následujícím okamžiku malé délky Δt, je přibližně $\lambda(t).\Delta t$:

$$P(t<X\leq t+\Delta t|X>t)\approx\frac{f(t)}{1-F(t)}\Delta t = \lambda(t).\Delta t$$

Intenzita poruch má pro většinu výrobků z technické praxe charakteristický tvar **vanové křivky**.

Jedním ze základních spojitých rozdělení pravděpodobnosti je rozdělení rovnoměrné (rektangulární) na intervalu (a;b).

Název rozdělení	Popis	Hustota pravděpodobnosti	EX	DX
Rovnoměrné na (a;b)	f(x) je na (a;b) konstantní, jinde nulová	$f(x)=\begin{cases}\frac{1}{b-a} & x\in\langle a;b\rangle\\ 0 & jinde\end{cases}$	$\frac{a+b}{2}$	$\frac{(a-b)^2}{12}$

Následující tři rozdělení jsou založena na Poissonovském procesu, tj. na předpokladu, že jednotlivé události nastávají nezávisle na sobě, s konstantní rychlostí výskytu. Tato rozdělení se používají většinou pro popis náhodné veličiny definované jako doba do k-té události (poruchy), popř. doba mezi událostmi (poruchami).

Název rozdělení	Popis	Hustota pravděpodobnosti, Distribuční funkce, intenzita poruch	EX	DX
Exponenciální	doba do první události, doba mezi událostmi (popisuje pouze období stabilního života)	$f(t)=\lambda\cdot e^{-\lambda t};\quad t>0;\lambda>0$ $F(t)=1-e^{-\lambda t};\quad t>0;\lambda>0$ $\lambda(t)=\lambda=konst.;\quad t>0;\lambda>0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Erlangovo	doba do k-té události	$f(t)=\lambda\cdot e^{-\lambda t}\cdot\frac{(\lambda t)^{k-1}}{(k-1)!};\quad t>0$ $F(t)=1-e^{-\lambda t}\cdot\sum_{j=0}^{k-1}\frac{(\lambda t)^j}{j!}$ $\lambda(t)=\frac{\lambda}{(k-1)!\sum_{j=0}^{k-1}\frac{1}{(k-1-j)!(\lambda t)^j}}$	$\frac{k}{\lambda}$	$\frac{k}{\lambda^2}$

Weibullovo	doba do první události (poruchy) (vhodná volba β umožňuje použití v libovolném období intenzity poruch)	$f(t) = \frac{\beta}{\Theta}\left(\frac{t}{\Theta}\right)^{\beta-1} e^{-\left(\frac{t}{\Theta}\right)^{\beta}}$ $F(t) = 1 - e^{-\left(\frac{t}{\Theta}\right)^{\beta}}$ $\lambda(t) = \frac{\beta}{\Theta}\cdot\left(\frac{t}{\Theta}\right)^{\beta-1}$ $t > 0;\ \Theta > 0;\ \beta > 0$		

Nejdůležitějším pravděpodobnostním rozdělením popisujícím chování velkého množství náhodných jevů v technice, ekonomii i v přírodních vědách je rozdělení normální, jehož parametry jsou střední hodnota μ a rozptyl σ^2, a jeho speciální typ rozdělení normované normální s parametry $\mu=0$ a $\sigma^2=1$.

Název rozdělení	**Vlastnosti**	**Hustota pravděpodobnosti, Distribuční funkce**	**EX**	**DX**
Normované normální	distribuční funkce $\Phi(z)$ je tabelovaná, hustota pravděpodobnosti je sudá funkce („Gaussův klobouk“)	$\varphi(x) = \frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}; \qquad -\infty < x < \infty$ $\Phi(x) = \frac{1}{\sqrt{2\pi}}\cdot\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$	0	1
Normální	distribuční funkci určujeme pomocí standardizace normální náhodné veličiny $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2}; \qquad -\infty < x < \infty$ $F(x) = \frac{1}{\sigma\sqrt{2\pi}}\cdot\int_{-\infty}^{x} e^{-\left(\frac{t-\mu}{\sqrt{2}\sigma}\right)^2} dt$	μ	σ^2

V SPC (spolehlivost a jakost, statistická kontrola jakosti) se pak velmi často používá **metoda 6 sigma**.

Při popisu náhodných veličin nabývajících pouze kladných hodnot a to zejména v případech, kdy hustota pravděpodobnosti je asymetrická používáme logaritmicko-normální rozdělení.

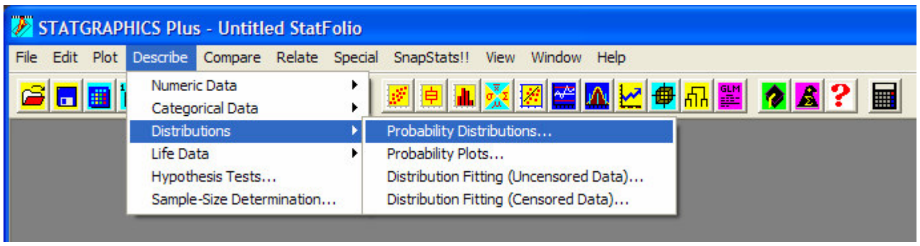
Název rozdělení	**Vlastnosti**	**Hustota pravděpodobnosti**	**EX**	**DX**
Logaritmicko-normální	distribuční funkci určujeme převodem na distribuční funkci normovaného normálního rozdělení $F(x) = \begin{cases} \Phi\left(\frac{\ln x - \mu}{\sigma}\right); & \text{pro } x > 0 \\ 0 & \text{pro } x \le 0 \end{cases}$	$f(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}}\cdot e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}; & \text{pro x} > 0 \\ 0 & \text{pro x} \le 0 \end{cases}$	$e^{\mu+\frac{\sigma^2}{2}}$	$e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right)$

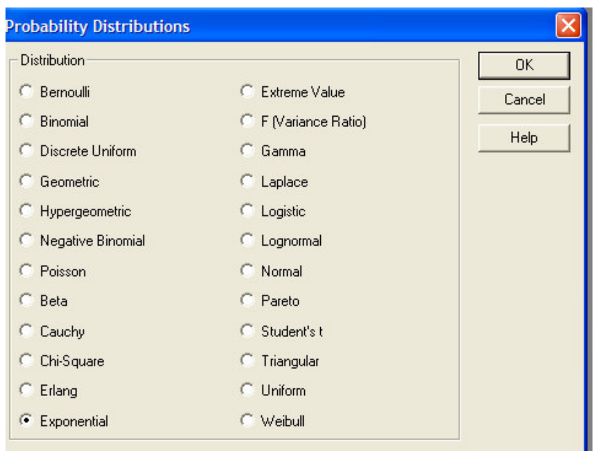
7.1. Seznamte se s možnostmi zpracování spojité náhodné veličiny ve Statgraphicsu.

Ukážeme si jaké **informace nám Statgraphics nabízí pro spojité náhodné veličiny** (konkrétně pro exponenciální).

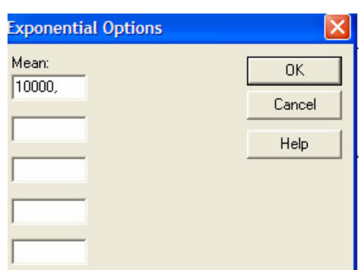
Obdobně jako v případě diskrétního rozdělení volíme:

Menu Describe\Distributions\Probability Distributions a **v okně Probability Distributions** zaškrtneme typ rozdělení (Exponential).





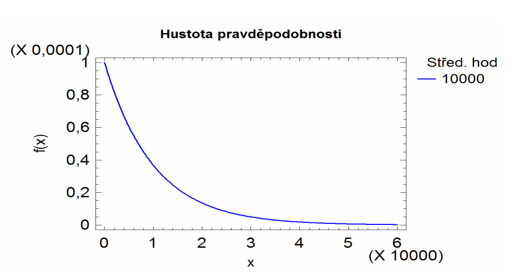
Provedeme RC na textový výstup (levé dolní okno) a v menu **Analysis Options** nastavíme parametry daného rozdělení. (V případě exponenciálního rozdělení zadáváme jako parametr střední hodnotu $\left(\frac{1}{\lambda}\right)$).

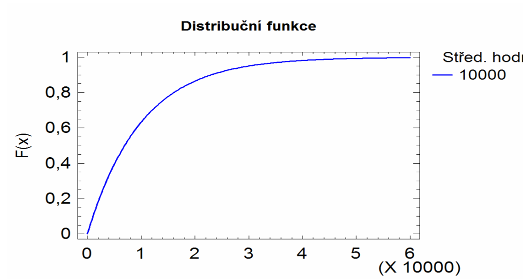


Všimněte si, že Statgraphics umožňuje současně studovat informace o 1 až 5 různými náhodnými veličinami daného typu rozdělení.

Grafické výstupy:

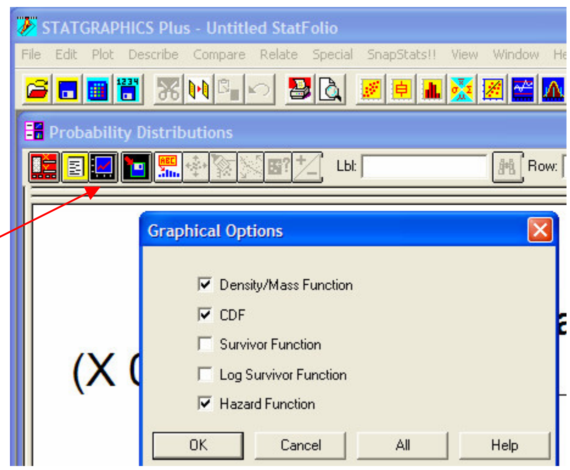
V pravém horním rohu se zobrazí **hustota pravděpodobnosti** (daného (daných) rozdělení, v pravém dolním rohu najdeme **distribuční funkci** tohoto (těchto) rozdělení. Nastavení grafických parametrů obou grafů provádíme přes **menu Graphics Options** (zobrazí se po provedení RC na oblast příslušného grafu.)

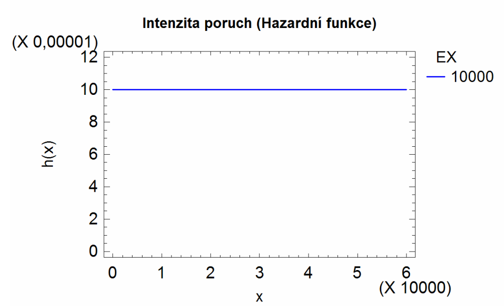




Další funkcí, která popisuje spojitou náhodnou veličinu je **intenzita poruch (hazardní funkce)**. Její graf získáme tak, že klikneme na **ikonu Graphical Options** a zaškrtneme **Hazard Function**.

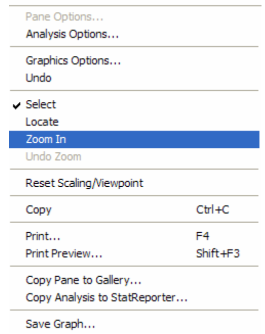
Graphical Options



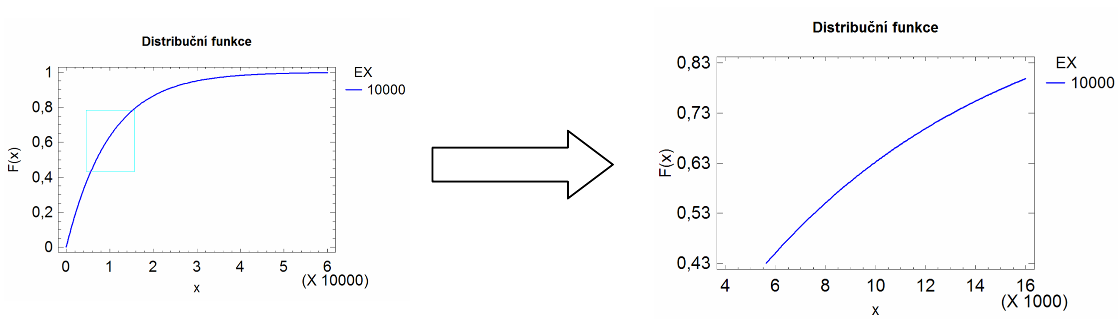


Ve všech případech grafického výstupu můžeme používat funkci **Zoom**.

Provedeme RC na oblast příslušného grafu a zvolíme položku **Zoom In**.

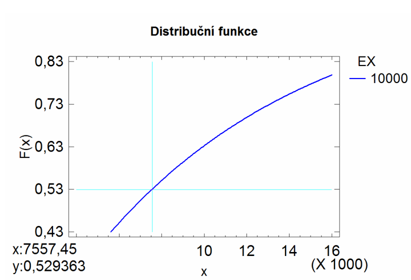


Obecně známým způsobem nyní můžeme zobrazit výřez grafu.



Statgraphics nám rovněž **umožňuje pomocí osového kříže odečítat hodnoty na grafu**.

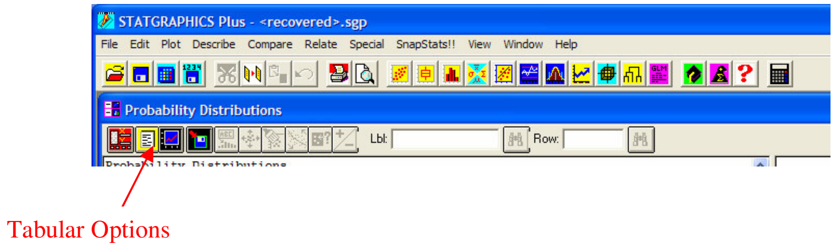
Provedeme RC na oblast příslušného grafu a zvolíme položku **Locate**. Na grafu se objeví osový kříž a souřadnice středu tohoto kříže. Myší můžeme osovým křížem pohybovat a zároveň odečítat souřadnice bodu.

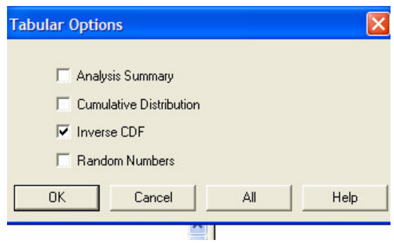


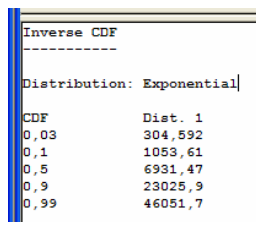
Textový výstup:

V levé dolním rohu najdeme v textovém výstupu hodnoty **distribuční funkce** ($P(X < x)$), **hustoty pravděpodobnosti** f(x) a hodnotu **doplňku distribuční funkce** ($P(X > x)$). To vše pro x=0. Jak již víme, konkrétní hodnotu, v níž chceme uvedené funkce určit, nastavíme v menu **Pane Options** (RC na oblast levého dolního okna).

Statgraphics nám dále umožňuje najít **hodnoty kvantilů náhodných veličin**. Máme-li zobrazeno rozdělení náhodné veličiny, jejíž kvantily nás zajímají, stačí klinout na **ikonu Tabular Options** (žlutá ikona) a zvolit **Inverse CDF**. (Všichni přece víme, že kvantily jsou inverzní funkcí k funkci distribuční ☺.)





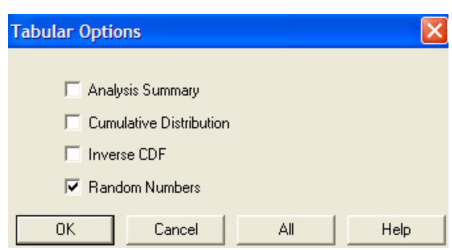


```
Inverse CDF
-----------

Distribution: Exponential

CDF           Dist. 1
0,03          304,592
0,1           1053,61
0,5           6931,47
0,9           23025,9
0,99          46051,7
```

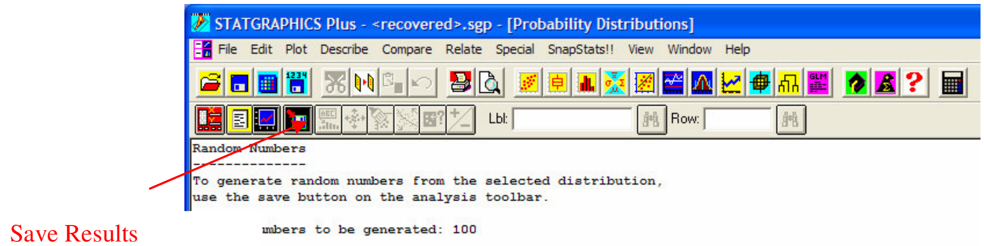
Hodnoty vybraných kvantilů se nám zobrazí v textovém výstupu. Jejich změnu (nastavení požadovaného p) provedeme v okně **Pane Options** (RC na příslušný textový výstup).

Na závěr seznámení se zpracováním spojité náhodné veličiny si ukážeme **jak ve Statgraphicsu generovat náhodná čísla podléhající určitému rozdělení**.

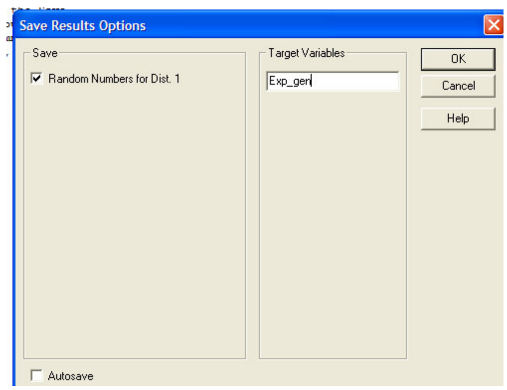
Máme-li zobrazeno rozdělení příslušné náhodné veličiny, klikneme na ikonu **Tabular Options** (žlutá ikona) a zvolíme položku **Random Numbers**.



V textovém výstupu vztahujícímu se k této problematice najdeme informaci o tom, že pro vygenerování 100 náhodných čísel podléhajících danému rozdělení máme kliknout na **ikonu Save Results** (ikona na níž je zobrazena disketa s červenou šipkou)



V okně **Save results Options** zaškrtneme požadavek na generování dat s požadovaným rozdělením a zvolíme název pro tato data. Po odsouhlasení (OK) najdeme generována data pod zvoleným názvem v tabulce (datový vstup), kterou právě používáme.



7.2. Výrobce žárovky XX ví, že průměrná životnost žárovek XX je 10.000 h. V rámci své propagační kampaně chce garantovat dobu T, do níž se nespálí více než 3% žárovek. Určete tuto dobu.

Řešení:

X ... životnost žárovky (doba do poruchy) má exponenciální rozdělení

$$X \to E(\lambda)$$

- Určíme parametr λ:

$$\begin{aligned} EX &= \frac{1}{\lambda} \\ EX &= 10.000\, h \end{aligned} \quad \Rightarrow \quad \lambda = 10^{-4}\, h^{-1}$$

- Na základě zadané pravděpodobnosti najdeme dobu T:

$$\begin{aligned} P(X < T) &\le 0{,}03 \\ F(T) &\le 0{,}03 \\ 1 - e^{-\lambda T} &\le 0{,}03 \\ 0{,}97 &\le e^{-\lambda T} \end{aligned}$$

$$-\frac{\ln(0{,}97)}{\lambda} \geq T$$

$$T \leq -10^4 \cdot \ln(0{,}97) \Rightarrow \underline{\underline{T \cong 304\ h}}$$

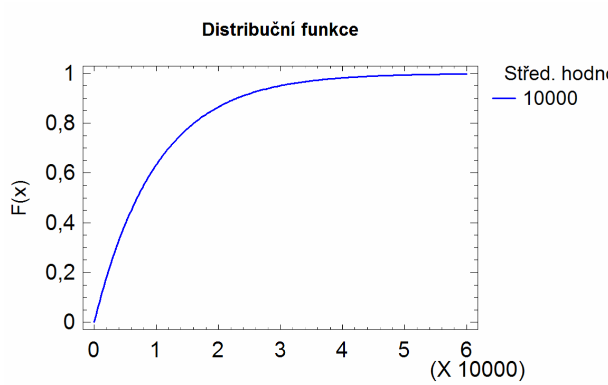
Výrobce může tvrdit, že více než 97% žárovek má životnost delší než 304 hodin.

Řešení ve Statgraphicsu:

Postup řešení zůstává naprosto stejný, pouze k výpočtu použijeme Statgraphics, nikoliv kalkulačku.

Statgraphics použijeme ve chvíli, kdy máme řešit nerovnici: $F(T) \quad \leq 0{,}03$

Zobrazíme si výstupy pro $X \to E(0{,}0001)$. Distribuční funkce spojité náhodné veličiny je funkce rostoucí a proto uvedenou nerovnici můžeme upravit: $T \quad \leq F^{-1}(0{,}03)$



Potřebujeme tedy najít 3% ní kvantil náhodné veličiny X. Postup při hledání kvantilů je popsán v příkladě 7.1. (ikona Tabular Options)

```
Inverse CDF
-----------

Distribution: Exponential

CDF            Dist. 1
0,03           304,592
0,1            1053,61
0,5            6931,47
0,9            23025,9
0,99           46051,7
```

Řešení odečteme z textového výstupu: $\underline{\underline{T \quad \leq 304{,}6}}$

7.3. Předpokládejme, že doba do poruchy určitého systému je modelována Weibullovým rozdělením s lineárně rostoucí intenzitou poruch. (Θ =50)

a) Jaká je intenzita poruch systému po deseti hodinách funkce?
b) Jaká je pravděpodobnost, že systém bude pracovat bez poruchy během počátečních 100 hodin?

Řešení:

$$X \text{ ... doba do poruchy, } (X \to W(50;\beta))$$

Hodnotu parametru β určíme na základě poznámky, že intenzita poruch je lineárně rostoucí. Obecný tvar intenzity poruch Weibullova rozdělení je:

$$\lambda(t) = \frac{\beta}{\Theta}.\left(\frac{t}{\Theta}\right)^{\beta-1}; \qquad t > 0;\ \Theta > 0;\ \beta > 0$$

z čehož vyplývá, že β = 2.

$$X \to W(50;2)$$

ada) Hledanou intenzitu poruch určíme dosazením do obecného vztahu:

$$\underline{\underline{\lambda(10)}} = \frac{2}{50}.\left(\frac{10}{50}\right)^{2-1} \underline{\underline{= 0{,}008}}$$

Intenzita poruch daného systému je po 10 hodinách provozu 0,008. Tj. pokud byl systém po 10 hodin bezporuchový, pak pravděpodobnost, že v následujícím velmi krátkém časovém intervalu Δt dojde k poruše, je 0,008 .Δt.

adb) Pravděpodobnost, že systém bude prvních 100 hodin bezporuchový určíme přes jev opačný, jehož pravděpodobnost udává distribuční funkce.

$$F(t) = 1 - e^{-\left(\frac{t}{\Theta}\right)^{\beta}}; \qquad t > 0;\ \Theta > 0;\ \beta > 0$$

$$\underline{\underline{P(X > 100)}} = 1 - F(100) = 1 - \left[1 - e^{-\left(\frac{100}{50}\right)^2}\right] = e^{-\left(\frac{100}{50}\right)^2} = e^{-4} \underline{\underline{= 0{,}018}}$$

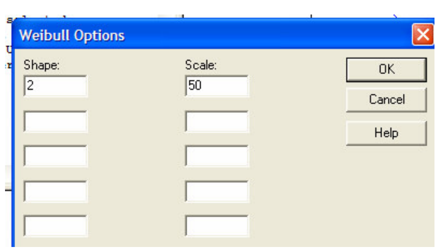
Pravděpodobnost, že daný systém bude prvních 100 hodin bezporuchový je 1,8%.

Řešení ve Statgraphicsu:

Začneme opět tím, že si dané rozdělení zobrazíme.

$$X \to W(50;2)$$

Při nastavování parametru Weibullova rozdělení musíte vědět, že **Θ je parametr měřítka (scale)** a **β je parametr tvaru (shape)**.



ada)

Nabízí se nám dvě možnosti řešení:

1.) $\lambda(t)=\dfrac{f(t)}{1-F(t)}$, stačí tedy určit hodnotu hustoty pravděpodobnosti a hodnotu distribuční funkce v příslušném bodě (10).

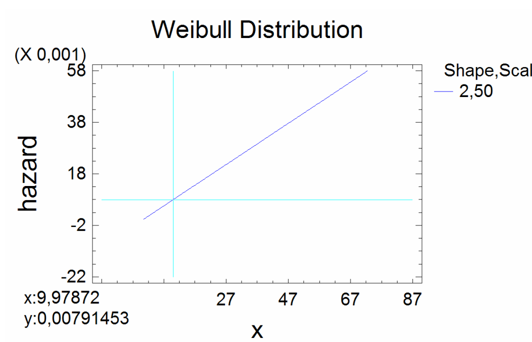
Postup, jak tyto údaje získat již nebudeme opakovat. Nevíte-li, vraťte se k příkladu 7.1.

```
Distribution: Weibull

              Lower Tail Area (<)
Variable      Dist. 1       Dist. 2
10            0,0392106

              Probability Density
Variable      Dist. 1       Dist. 2
10            0,00768632

              Upper Tail Area (>)
Variable      Dist. 1       Dist. 2
10            0,960789
```

$$\underline{\underline{\lambda(10)}}=\frac{f(10)}{1-F(10)}=\frac{0{,}00768632}{1-0{,}0392106}\cong\underline{\underline{0{,}008}}$$

2.) Můžeme zkusit odečíst požadovanou hodnotu z grafu intenzity poruch:

(Postup je uveden opět v příkladu 7.1.)



adb) Hledanou pravděpodobnost odečteme přímo z textového výstupu pro danou náhodnou veličinu, pro x = 100.

```
Distribution: Weibull

              Lower Tail Area (<)
Variable      Dist. 1       Dist. 2
100           0,981684

              Probability Density
Variable      Dist. 1       Dist. 2
100           0,00146525

              Upper Tail Area (>)
Variable      Dist. 1       Dist. 2
100           0,0183156
```

$$\underline{\underline{P(X>100)}}\cong\underline{\underline{0{,}018}}$$

7.4. Určete:
a) $\Phi(0{,}54)$
b) $\Phi(-2{,}42)$
c) $z_{0,75}$
d) $z_{0,25}$

Řešení:

ada) Příslušnou distribuční funkci nalezneme v Tabulce 1:

V prvním sloupci je uveden argument distribuční funkce s přesností na jedno desetinné místo (0,5), identifikátor druhého sloupce udává druhé desetinné místo argumentu (4).

$$\underline{\underline{\Phi(0{,}54) = 0{,}705}}$$

adb) Pro nalezení distribuční funkce záporného argumentu musíme použít převodní vztah:

$$\Phi(z) = 1 - \Phi(-z); \qquad -\infty < z < \infty$$

V našem případě:

$$\Phi(-2{,}42) = 1 - \Phi(2{,}42)$$
$$\Phi(-2{,}42) = 1 - 0{,}992$$
$$\underline{\underline{\Phi(-2{,}42) = 0{,}008}}$$

adc) Pro určení 100p%-ního kvantilu se musíme pokusit najít p v jádru tabulky a určit pro ně příslušnou hodnotu z_p.

$$\Phi(z_p) = p$$

V našem případě:

$$\Phi(z_{0,75}) = 0{,}75$$
$$\underline{\underline{z_{0,75} \cong 0{,}67}}$$

add) V Tabulce 1 nalezneme hodnoty (50 až 100)%-ních kvantilů. Pro nalezení (0 až 50)%-ních kvantilů musíme použít převodní vztah mezi kvantily, který si tímto odvodíme:

$$\Phi(z_p) = p; \qquad \Phi(z_{1-p}) = 1 - p$$
$$1 - \Phi(z_p) = 1 - p$$
$$\Phi(-z_p) = \Phi(z_{1-p})$$
$$-z_p = z_{1-p}$$

V našem případě:

- $z_{0,25}$ v Tabulce 1 nenalezneme.
- $z_{0,25} = -z_{1-0,25} = -z_{0,75}$
- Nalezneme $z_{0,75}$:

$$\Phi(z_{0,75}) = 0,75$$
$$z_{0,75} = 0,67$$

- Určíme $z_{0,25}$: $\underline{\underline{z_{0,25}}} = -z_{0,75} \underline{\underline{= -0,67}}$

Řešení ve Statgraphicsu:

Jde o určení hodnot distribuční funkce a hodnot kvantilů náhodné veličiny s normovaným normálním rozdělením.

Opět začneme tím, že si zobrazíme rozdělení této náhodné veličiny ($X \rightarrow N(0;1)$):

Zvolíme si menu: **Describe\Distributions\Probability Distributions** a vybereme rozdělení normální (**Normal**). Výstupem je popis normované normální náhodné veličiny (μ=0; σ^2=1), proto nemusíme parametry rozdělení měnit.

ada, adb)
Hodnoty distribuční funkce určíme tak, že v menu Pane Options (RC na textový výstup) nastavíme požadované x:

```
Distribution: Normal

              Lower Tail Area (<)
Variable      Dist. 1        Dist. 2
0,54          0,705403
-2,42         0,00776023

              Probability Density
Variable      Dist. 1        Dist. 2
0,54          0,344818
-2,42         0,0213407

              Upper Tail Area (>)
Variable      Dist. 1        Dist. 2
0,54          0,294597
-2,42         0,99224
```

Je zřejmé, že:

$$\underline{\underline{\Phi(0,54) \cong 0,705}}$$
$$\underline{\underline{\Phi(-2,42) \cong 0,008}}$$

adc, add)
Hodnoty kvantilů získáme pomocí ikony Tabular Options (zaškrtneme Inverse CDF). Pomocí menu Pane Options navolíme požadované p.

```
-----------

Distribution: Normal

CDF           Dist. 1
0,25          -0,674492
0,75          0,674492
0,5           0
0,9           1,28155
0,99          2,32635
```

Odečteme:

$$\underline{\underline{z_{0,25} \cong 0,674}}$$
$$\underline{\underline{z_{0,75} \cong -0,674}}$$

7.5. Nechť náhodná veličina X má normální rozdělení se střední hodnotou 10 a směrodatnou odchylkou 5. Určete:
a) F(7)
b) $x_{0,75}$
c) $x_{0,30}$

Řešení:

$$X \to N(10;25) \Rightarrow \mu = 10;\ \sigma^2 = 25$$

ada) Distribuční funkci normální náhodné veličiny určíme pomocí standardizace:

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$
$$F(7) = \Phi\left(\frac{7-10}{\sqrt{25}}\right) = \Phi(-0,6)$$
$$F(7) = 1 - \Phi(0,6)$$
$$F(7) = 1 - 0,726 \quad (viz.\ Tabulka\ 1)$$
$$\underline{\underline{F(7) = 0,274}}$$

adb) Postup při určení horního kvartilu je následující (opět využijeme standardizace):

$$F(x_{0,75}) = 0,75$$
$$\Phi\left(\frac{x_{0,75}-10}{\sqrt{25}}\right) = 0,75$$
$$\frac{x_{0,75}-10}{\sqrt{25}} = 0,67 \qquad (viz.\ Tabulka\ 1)$$
$$x_{0,75} = 5 \cdot 0,67 + 10$$
$$\underline{\underline{x_{0,75} = 13,35}}$$

adc) Poněkud odlišný postup musíme použít pro nalezení 30%-ního kvantilu:

$$F(x_{0,30}) = 0,30$$
$$\Phi\left(\frac{x_{0,30}-10}{\sqrt{25}}\right) = 0,30$$

V této fázi však ještě nemůžeme použít Tabulku 1, protože v jádru tabulky se nacházejí pouze hodnoty (0,50 až 1,00). A proto rovnici upravíme do vhodnějšího tvaru:

$$\Phi\left(\frac{x_{0,30}-10}{\sqrt{25}}\right) = 0{,}30$$

$$1-\Phi\left(\frac{x_{0,30}-10}{\sqrt{25}}\right) = 1-0{,}30$$

$$\Phi\left(-\frac{x_{0,30}-10}{\sqrt{25}}\right) = 0{,}70$$

A nyní již tabulky můžeme použít:

$$\Phi\left(-\frac{x_{0,30}-10}{\sqrt{25}}\right) = 0{,}70$$

$$-\frac{x_{0,30}-10}{\sqrt{25}} = 0{,}525 \qquad (viz.\,Tabulka\;1)$$

$$x_{0,30} = -5\cdot 0{,}525+10$$

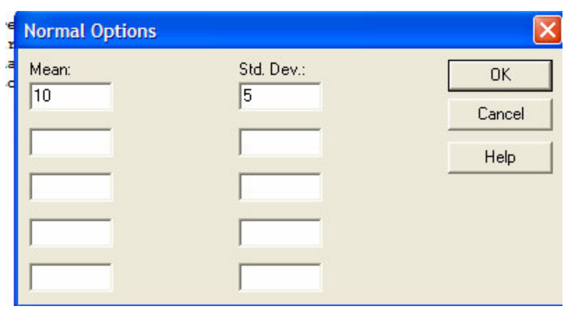
$$\underline{\underline{x_{0,30} = 7{,}375}}$$

Řešení ve Statgraphicsu:

Jde o určení hodnoty distribuční funkce a hodnot kvantilů náhodné veličiny s normálním rozdělením, $X \to N(10;25) \Rightarrow \mu = 10; \sigma^2 = 25$.

Zobrazíme si rozdělení této náhodné veličiny. Zvolíme si menu: **Describe\Distributions\Probability Distributions** a vybereme rozdělení normální (**Normal**).

Výstupem je popis normované normální náhodné veličiny ($\mu=0$; $\sigma^2=1$), proto musíme parametry rozdělení změnit v **Analysis Options**.



Všimněme si, že jako parametry tohoto rozdělení nastavujeme střední hodnotu a směrodatnou odchylku (nikoliv rozptyl).

ada)
Hodnotu distribuční funkce určíme tak, že v menu Pane Options (RC na textový výstup) nastavíme požadované x:

Je zřejmé, že:

```
Distribution: Normal

              Lower Tail Area (<)
Variable      Dist. 1       Dist. 2
7             0,274252

              Probability Density
Variable      Dist. 1       Dist. 2
7             0,0666449

              Upper Tail Area (>)
Variable      Dist. 1       Dist. 2
7             0,725748
```

$$\underline{\underline{F(7) \quad \cong 0{,}274}}$$

adb, adc)
Hodnoty kvantilů získáme pomocí ikony Tabular Options (zaškrtneme Inverse CDF). Pomocí menu Pane Options navolíme požadované p.

```
Distribution: Normal

CDF           Dist. 1
0,75          13,3725
0,3           7,37799
0,5           10
0,9           16,4078
0,99          21,6318
```

Odečteme:

$$\underline{\underline{x_{0,75} \cong 13{,}37}}$$

$$\underline{\underline{x_{0,30} \cong 7{,}38}}$$

7.6. Stanovme pravděpodobnost, že náhodná veličina X mající rozdělení $N(\mu,\sigma^2)$ nabude hodnoty z intervalu $(\mu - k\cdot\sigma;\, \mu + k\cdot\sigma)$ pro dané kladné k.

Řešení:

Pro k>0:

$$\underline{\underline{P(\mu - k\sigma < X < \mu + k\sigma)}} = F(\mu + k\sigma) - F(\mu - k\sigma) = \Phi\left(\frac{(\mu + k\sigma) - \mu}{\sigma}\right) - \Phi\left(\frac{(\mu - k\sigma) - \mu}{\sigma}\right) =$$

$$= \Phi(k) - \Phi(-k) = \Phi(k) - [1 - \Phi(k)] = \underline{\underline{2\cdot\Phi(k) - 1}}$$

Následující tabulka uvádí hodnoty této pravděpodobnosti pro některé hodnoty k:

k	$P(\mu - k\sigma < \mathrm{X} < \mu + k\sigma)$
1	0,683
1,64	0,900
1,96	0,950
2,58	0,990
3	0,998

7.7. Firma získá z každého prodaného výrobku 100,-Kč. Za výměnu během záruční lhůty zaplatí 300,-Kč. Životnost výrobku v letech má normální rozdělení N(3;1). Jakou záruční dobu v měsících má firma stanovit, aby střední (průměrný) zisk byl alespoň 60,-Kč/výrobek?

Řešení:

X … počet reklamovaných výrobků (z jednoho prodaného)
Y … zisk z jednoho prodaného výrobku
Z … životnost výrobku
$\mathrm{T_Z}$ … záruční doba

Je zřejmé, že X má alternativní rozdělení, jehož parametr p je roven pravděpodobnosti, že dojde k reklamaci výrobku během záruční doby:

$$X \to A(p); \quad p = P(Z < \mathrm{T_Z})$$

Zisk z jednoho výrobku (Y) je dán jako:

$$Y = 100 - 300 \cdot X$$

A tedy: $EY = 100 - 300 \cdot EX$

$EX = p = P(Z < T_Z)$, kde $Z \to N(3;1)$ a proto $EX = F(T_Z) = \Phi\left(\frac{T_Z - 3}{1}\right) = \Phi(T_Z - 3)$

Nyní stačí vyřešit nerovnici popisující požadavek na záruční dobu:

$$\begin{aligned}
EY &\geq 60 \\
100 - 300 \cdot EX &\geq 60 \\
100 - 300 \cdot \Phi(T_Z - 3) &\geq 60 \\
\Phi(T_Z - 3) &\leq \frac{4}{30} \\
\Phi(T_Z - 3) &\leq 0{,}133
\end{aligned}$$

Řešení této nerovnice nelze najít v tabulkách, proto nerovnici upravíme:

$$\begin{aligned}
1 - \Phi(T_Z - 3) &\geq 1 - 0{,}133 \\
\Phi(-(T_Z - 3)) &\geq 0{,}867
\end{aligned}$$

viz. Tabulka 1.:

$$\begin{aligned}
-(T_Z - 3) &\geq 1{,}11 \\
T_Z &\leq 1{,}89\ let \\
T_Z &\leq 22{,}68\ m\check{e}\check{e}.
\end{aligned}$$

Firma by měla stanovit záruční dobu na 22 měsíců.

7.8. Nechť X je náhodná veličina s logaritmicko-normálním rozdělením s parametry: μ=2; σ^2=9. Určete:

a) pravděpodobnost, že náhodná veličina X je z intervalu (0;30)
b) medián daného rozdělení
c) střední hodnotu a rozptyl náhodné veličiny X

Řešení:

$$X \to LN(2;9)$$

ada) Pravděpodobnost, že náhodná veličina X je z intervalu (0;30) můžeme určovat rovněž jako pravděpodobnost, že náhodná veličina X je menší než 30, neboť log.-normální náhodná veličina může nabývat pouze kladných hodnot.

Připomeňme si postup při určování distribuční funkce log.-normální náhodné veličiny:

$$F(x) = \begin{cases} \Phi\left(\dfrac{\ln x - \mu}{\sigma}\right); & \text{pro } x > 0 \\ 0 & \text{pro } x \le 0 \end{cases}$$

A nyní již přejděme k určení hledané pravděpodobnosti:

$$\underline{\underline{P(0 < X < 30)}} = F(30) - F(0) = \Phi\left(\frac{\ln 30 - 2}{\sqrt{9}}\right) - 0 = \Phi(0{,}47) \underline{\underline{= 0{,}681}}$$

nebo

$$\underline{\underline{P(0 < X < 30)}} = P(X < 30) = F(30) = \Phi\left(\frac{\ln 30 - 2}{\sqrt{9}}\right) = \Phi(0{,}47) \underline{\underline{= 0{,}681}}$$

adb) Pro určení mediánu můžeme použít vztah pro 100p%-ní kvantil, který byl odvozen v Průvodci studiem:

$$x_p = e^{\mu + \sigma \cdot z_p}$$

$$z_{0,5} = 0 \quad (viz. Tabulka\ 1) \qquad \Rightarrow \quad \underline{\underline{x_{0,5}}} = e^{2+\sqrt{9}\cdot 0} = \underline{\underline{e^2 \cong 7{,}4}}$$

adc) Střední hodnotu a rozptyl určíme na základě výše uvedených vztahů:

$$EX = e^{\mu + \frac{\sigma^2}{2}} \Rightarrow \quad \underline{\underline{EX}} = e^{2+\frac{9}{2}} = \underline{\underline{e^{\frac{13}{2}} \cong 665{,}1}}$$

$$DX = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right) \Rightarrow \quad \underline{\underline{DX}} = e^{2\cdot 2+9}\left(e^9 - 1\right) \underline{\underline{\cong 3{,}6 \cdot 10^9}}$$

Řesení ve Statgraphicsu:

Pozor!!! Opět se setkáte s nestandardním přístupem ve Statgraphicsu. Parametry lognormálního rozdělení ve Statgraphicsu nejsou střední hodnota μ a rozptyl σ^2 (resp. směrodatná odchylka) příslušné normální náhodné veličiny Y (Y= ln X), ale střední hodnota a směrodatná odchylka náhodné veličiny X ($EX = e^{\mu+\frac{\sigma^2}{2}}$, $DX = e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right)$).

Chceme-li tedy pomocí Statgraphicsu určovat pro náhodnou veličinu X, která má lognormální rozdělení s parametry μ a σ^2, určovat pravděpodobnosti výskytu X na nějakém intervalu, musíme si nejdříve určit parametry, které pro tuto náhodnou veličinu požaduje Statgraphics.

Mean: $EX = e^{\mu+\frac{\sigma^2}{2}}$

Std. Deviation: $\sigma_X = \sqrt{e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right)}$

V našem případě využijeme výsledky, které jsme získali v bodě c).

$EX = 665{,}1$; $\sigma_X = \sqrt{3{,}6\cdot 10^9} = 6\cdot 10^4$

Nyní můžeme již známým způsobem zobrazit výstupy pro náhodnou veličinu s daným rozdělením.

Menu Describe\Distributions\Probability Distributions …

Zaškrtneme **Lognormal** a v menu **Analysis Option** (RC na oblast textového výstupu) zadáme požadované parametry (viz. výše).

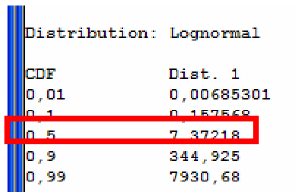
ada) Zobrazíme si hodnoty distribuční funkce v bodech 0 a 30 (**Pane Options**) a dosadíme:

```
-----------------------
Distribution: Lognormal

              Lower Tail Area (<)
Variable      Dist. 1       Dist. 2
0             0,0
30            0,680006

              Probability Density
Variable      Dist. 1       Dist. 2
0             0,0
30            0,00397245

              Upper Tail Area (>)
Variable      Dist. 1       Dist. 2
0             1,0
30            0,319994
```

$$\underline{\underline{P(0 < X < 30)}} = F(30) - F(0) = 0{,}680 - 0 = \underline{\underline{0{,}680}}$$

adb) Pro určení mediánu (obecně pro určení kteréhokoliv kvantilu) použijeme **ikonu Tabular Options** (žlutá ikona) a zaškrtneme **Inverse CDF**.

V generovaném výstupu přímo najdeme hodnotu mediánu (50%-ní kvantil). Pokud bychom chtěli získat hodnotu kvantilu, který nebyl automaticky vygenerován, provedeme RC na textový výstup a v **Pane Options** nastavíme požadované p.


Distribution: Lognormal

CDF	Dist. 1
0,01	0,00685301
0,1	0,157568
0,5	7,37218
0,9	344,925
0,99	7930,68

$$\underline{\underline{x_{0,5}}} = \cong \underline{\underline{7{,}37}}$$

7.9. Určete medián a 10%-ní kvantil náhodné veličiny s exponenciálním rozdělením se střední hodnotou 10s.

Řešení:

$$X \to Exp(\lambda)$$
$$EX = \frac{1}{\lambda} = 10\,s \quad \Rightarrow \quad \lambda = 0{,}1\,s^{-1}$$
$$F(t) = 1 - e^{-\lambda t}; \qquad \lambda > 0;\ \mathrm{t} \geq 0$$

Pro kvantily spojité náhodné veličiny platí: $F(x_p) = p$. Vzhledem k tomu, že musíme určit jak medián, tak i 10% ní kvantil, určíme si obecný vztah pro 100p% ní kvantil exponenciální NV:

$$F(x_p) = p$$
$$1 - e^{-\lambda \cdot x_p} = p$$
$$1 - p = e^{-\lambda \cdot x_{0,5}}$$
$$\ln(1-p) = \ln\left(e^{-\lambda \cdot x_p}\right)$$
$$\ln(1-p) = -\lambda \cdot x_p$$
$$\underline{\underline{x_p = \frac{-\ln(1-p)}{\lambda}}}$$

A nyní již stačí dosadit:

Medián $x_{0,5}$:

$$x_{0,5} = \frac{-\ln(1-0{,}5)}{\lambda} = \frac{-\ln\left(\frac{1}{2}\right)}{\lambda} = \frac{-\ln\left(2^{-1}\right)}{\lambda} = \frac{\ln 2}{\lambda}$$

Pro $X \to Exp(0{,}1)$: $\underline{\underline{x_{0,5}}} = 10 \cdot \ln 2 \underline{\underline{\cong 6{,}93}}$

$x_{0,1}$:

$$x_{0,1} = \frac{-\ln(1-0{,}1)}{\lambda} = \frac{-\ln(0{,}9)}{\lambda}$$

Pro $X \to Exp(0{,}1)$: $\underline{\underline{x_{0,1}}} = -10 \cdot \ln(0{,}9) \underline{\underline{\cong 1{,}05}}$

Řešení ve Statgraphicsu:

Zobrazíme příslušnou náhodnou veličinu.

Menu Describe\Distributions\Probability Distributions …

Zaškrtneme **Exponential**
V menu **Analysis Options** (RC na textový výstup) zadáme jako parametr exponenciální NV její střední hodnotu. (V našem případě nemusíme provádět, přednastavená střední hodnota je 10.)

Klikneme na ikonu **Tabular Options** (žlutá ikona) a zaškrtneme **Inverse CDF**, odečteme medián a 10%-ní kvantil.

Distribution: Exponential

CDF	Dist. 1
0,01	0,100503
0,1	1,05361
0,5	6,93147
0,9	23,0259
0,99	46,0517

$\underline{\underline{x_{0,5} \cong 6{,}93}}$

$\underline{\underline{x_{0,1} \cong 1{,}05}}$

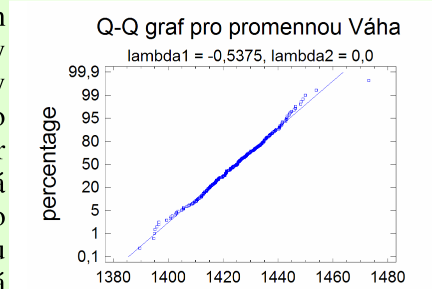
Nástroje ověření normality

Normalita je hlavním předpokladem o datech v drtivé většině analýz a testů (parametrické testy, Shewhartovy regulační diagramy, indexy způsobilosti…). Jde o předpoklad, že data pocházejí z normálního rozdělení. **Ověření normality je nezbytný krok před každou zodpovědnou analýzou jednorozměrných dat**.

Grafické znázornění a vizuální posouzení
(uživatel musí mít alespoň minimální znalosti o konstrukci a používání diagnostických exploratorních grafů). Nejčastěji se používá Q-Q graf, jádrové odhady hustoty, popř. kruhový graf.

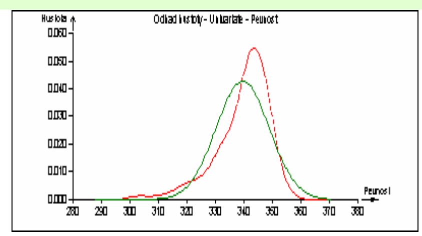
Q-Q graf

Jde o graf pro diagnostiku normality a odlehlých pozorování. Na ose x jsou vyneseny teoretické kvantily normálního rozdělení, na ose y jsou výběrové kvantily konstruované přímo z dat (viz. Exploratorní analýza). Pro normální data bez odlehlých pozorování má graf tvar přímky; pro normální data s odlehlými pozorovaními má tvar přímky s koncovými body ležícími mimo tuto přímku; pro systematicky zešikmená data s kladnou šikmostí (např. rozdělení lognormální, exponenciální) má nelineární konvexní tvar . Pro systematicky zešikmená data se zápornou šikmostí má nelineární konkávní tvar . Pro data s vyšší špičatostí než odpovídá normálnímu rozdělení, tedy s vysokou koncentrací dat kolem střední hodnoty (např. Laplaceovo rozdělení) má tvar konkávně-konvexní . Pro data s nižší špičatostí než odpovídá normálnímu rozdělení, tedy s malou koncentrací dat kolem střední hodnoty (např. rovnoměrné rozdělení) má tvar konvexně-konkávní . Proti statistikám má QQ-graf výhodu v možnosti vizuálně posoudit, zda je nelinearita způsobena jen několika body, nebo všemi daty.



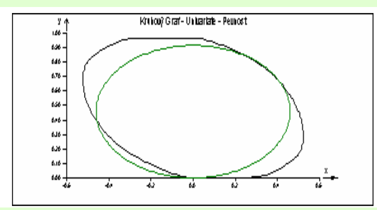
Odhad hustoty

Porovnání průběhu hustoty pravděpodobnosti normálního rozdělení (plná čára) s jádrovým odhadem hustoty vypočítaným na základě dat (přerušovaná čára). V případě normality a většího množství dat jsou si obě křivky blízké.

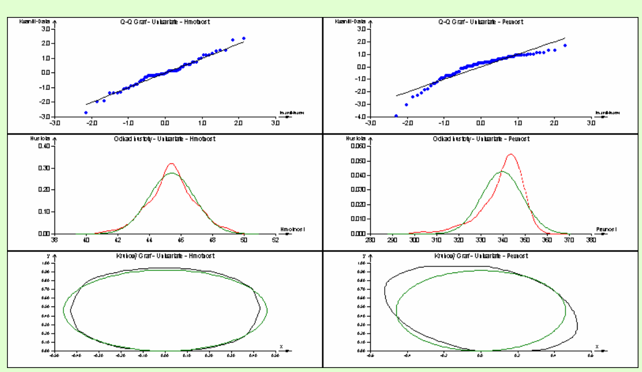


Kruhový graf

Slouží ke komplexnímu vizuálnímu posouzení normality na základě kombinace šikmosti a špičatosti. Zelený kruh (elipsa) je optimální tvar pro normální rozdělení, černý “kruh“ představuje data. V případě normálních dat se obě křivky téměř kryjí.



Ukázka výstupu (statistický software QC. Expert 2.5):



7.10. Vygenerujte ve Statgraphicsu náhodná čísla podléhající níže uvedeným rozdělením a otestujte jejich normalitu.

$NORM \rightarrow N(10;16)$

$EXP \rightarrow Exp(0{,}5)$

$ERLANG \rightarrow E(8;0.4)$

$WEIB \rightarrow W(50;8)$

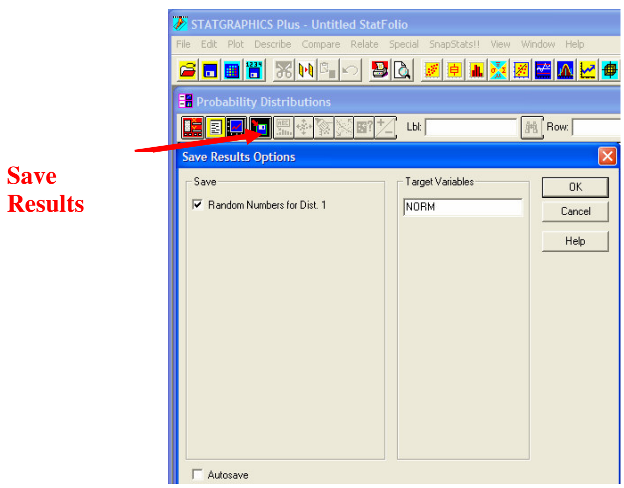
Řešení:

Jak již víme, generování náhodných čísel se provádí v menu:
Describe\Distributions\Probability Distributions …

Zaškrtneme požadovaný typ rozdělení a v menu **Analysis Options** (RC na oblast textového výstupu) nastavíme jeho parametry.

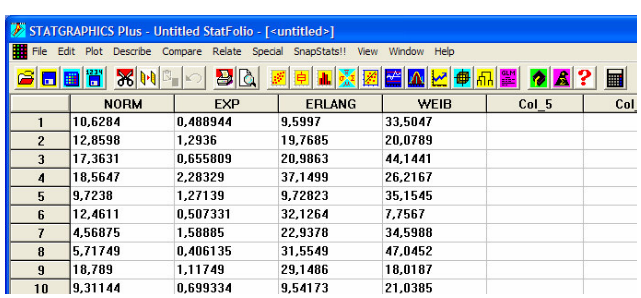
Proveďte pro NORM.

Nyní klikneme **na ikonu Save Results Options**, zaškrtneme Random Numbers for Dist. 1 a zadáme název pro požadovaná data (NORM).



Celý postup zopakujeme pro náhodná čísla podléhající exponenciálnímu, Erlangovu a Weibullovu rozdělení.

V datové tabulce bychom nyní měli najít 4 proměnné, přičemž pro každou z nich je generováno 100 hodnot.



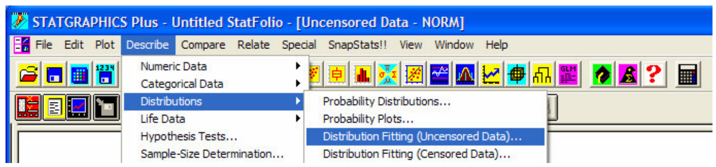
STATGRAPHICS Plus - Untitled StatFolio - [<untitled>]

	NORM	EXP	ERLANG	WEIB	Col_5	Col
1	10,6284	0,488944	9,5997	33,5047		
2	12,8598	1,2936	19,7685	20,0789		
3	17,3631	0,655809	20,9863	44,1441		
4	18,5647	2,28329	37,1499	26,2167		
5	9,7238	1,27139	9,72823	35,1545		
6	12,4611	0,507331	32,1264	7,7567		
7	4,56875	1,58885	22,9378	34,5988		
8	5,71749	0,406135	31,5549	47,0452		
9	18,789	1,11749	29,1486	18,0187		
10	9,31144	0,699334	9,54173	21,0385		

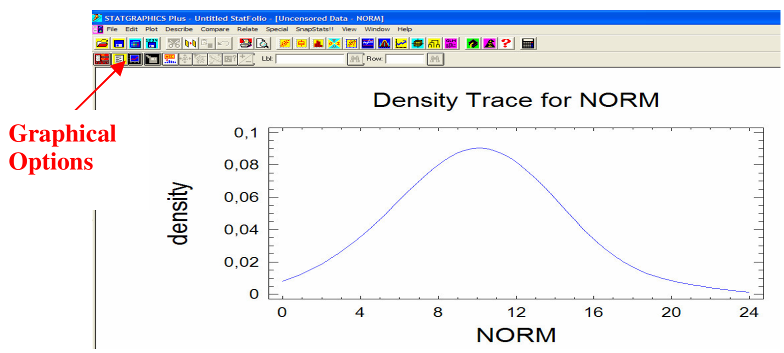
Nyní vyhodnotíme normalitu těchto dat.

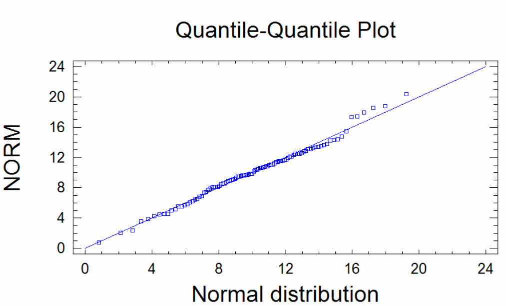
Co se týče grafů pro diagnostiku normality, Statgraphics nám nabízí pouze dva. A to empirickou (odhadovanou) hustotu pravděpodobnosti a Q-Q graf.

Oba tyto grafy získáme v menu **Describe\Distributions\Distributionn Fitting (Uncensored Data) …**



Jako Data zadáme testovanou proměnnou. Poté klikneme na **ikonu Graphical Options** a zaškrtneme požadované grafické výstupy (Density Trace a Quantille-Quantille graph).





Z grafu empirické hustoty pravděpodobnosti můžeme usuzovat na to, že jak šikmost, tak i špičatost rozdělení odpovídá normálnímu rozdělení. Rovněž Q-Q graf naznačuje, že tudovaná data můžeme považovat za výběr z normálního rozdělení (body leží v blízkosti vyznačené přímky).

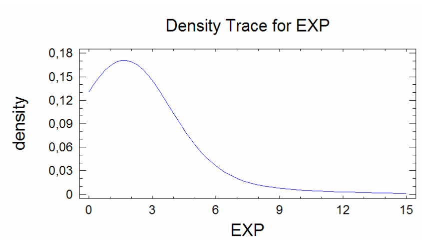
Co se týče šikmosti a špičatosti těchto dat, můžeme ji přesněji posoudit z číselných hodnot těchto statistik. Hodnoty šikmosti (skewness) a špičatosti (kurtosis) získáme v textovém výstupu menu **Describe\Numeric Dat\One-Variable Analysis …**, kde jako data zvolíme studovanou proměnnou (NORM).

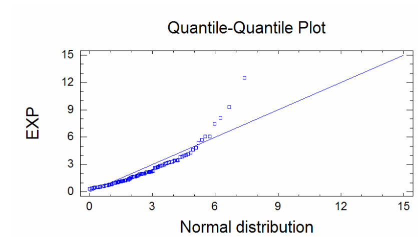
```
Summary Statistics for NORM

Count = 100
Average = 10,0385
Median = 9,99739
Variance = 13,9276
Standard deviation = 3,73197
Minimum = 0,78585
Maximum = 20,3545
Range = 19,5687
Lower quartile = 7,9604
Upper quartile = 12,3936
Stnd. skewness = 0,577646
Stnd. kurtosis = 0,670927
Coeff. of variation = 37,1765%
```

Vidíme, že hodnoty obou charakteristik jsou blízké nule – což rovněž svědčí ve prospěch normality dat.

Obdobné vyhodnocen ní provedeme pro zbylé 3 proměnné:

EXP:



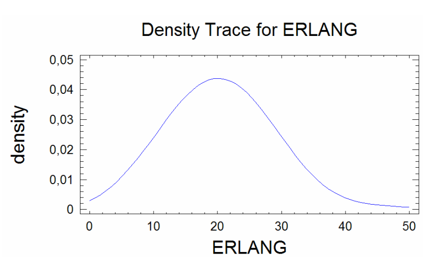


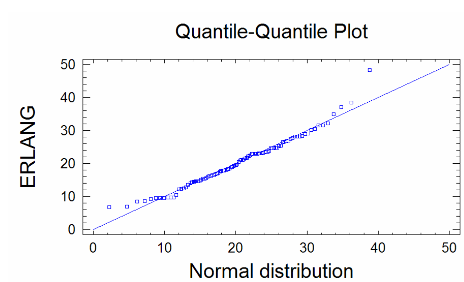
Z obou grafů je patrné, že data jsou pozitivně zešikmená (delší pravý chvost hustoty, resp. nelineární konvexní tvar Q-Q grafu)

```
Count = 100
Average = 2,20852
Median = 1,71475
Variance = 4,39839
Standard deviation = 2,09723
Minimum = 0,0324608
Maximum = 12,5153
Range = 12,4828
Lower quartile = 0,70285
Upper quartile = 3,09835
Stnd. skewness = 8,56102
Stnd. kurtosis = 12,9491
Coeff. of variation = 94,9612%
```

Hodnoty šikmosti (8,6) i špičatosti (12,9) ukazují na podstatný odklon od normality.

ERLANG:



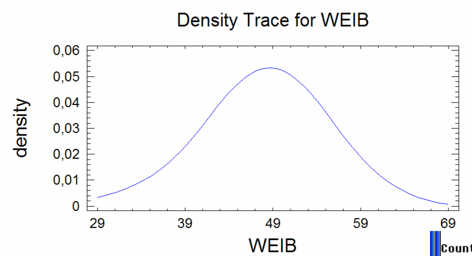


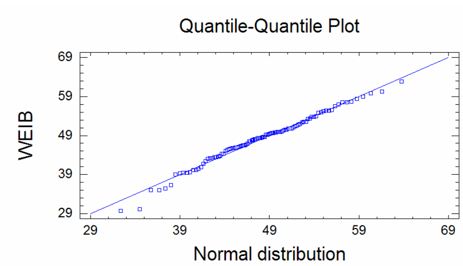
Z obou grafů je patrné, že data jsou mírně pozitivně zešikmená (delší pravý chvost hustoty, resp. nelineární konvexní tvar Q-Q grafu).

```
Count = 100
Average = 20,4852
Median = 20,7526
Variance = 55,0883
Standard deviation = 7,42215
Minimum = 6,80277
Maximum = 48,2561
Range = 41,4533
Lower quartile = 15,3885
Upper quartile = 24,7184
Stnd. skewness = 2,36655
Stnd. kurtosis = 2,18786
Coeff. of variation = 36,2317%
```

Naší domněnku potvrzují také hodnoty šikmosti i špičatosti. Na základě explorační statistiky můžeme data považovat za výběr podléhající normálnímu rozdělení.

WEIB:






```
Count = 100
Average = 48,0539
Median = 48,4961
Variance = 40,5989
Standard deviation = 6,37173
Minimum = 29,7229
Maximum = 62,7484
Range = 33,0255
Lower quartile = 44,3616
Upper quartile = 51,8838
Stnd. skewness = -1,5926
Stnd. kurtosis = 0,821499
Coeff. of variation = 13,2596%
```

Také tato data můžeme považovat za výběr z normálního rozdělení.

V kapitole testování hypotéz se naučíme vyhodnocovat normalitu dat na základě statistických testů.