10 ODHADY PARAMETRŮ NORMÁLNÍHO ROZDĚLENÍ

V praktických případech většinou nedokážeme přesně určit **parametry základního souboru** (populace). K jejich odhadu používáme charakteristiky příslušných výběrových souboru – **výběrové charakteristiky.**

Z metodického hlediska používáme dva typy odhadů parametrů:

bodový odhad, kdy parametr základního souboru aproximujeme jediným číslem
a
intervalový odhad (konfidenční interval), kdy tento parametr aproximujeme intervalem, v němž parametr leží s danou pravděpodobností. Této pravděpodobnosti říkáme **spolehlivost odhadu** a označujeme ji (1-α), α nazývame **hladinou významnosti**.

„Dobrý“ (věrohodný) odhad musí splňovat určité vlastnosti. Mezi základní vlastnosti věrohodných odhadů patří:

nestrannost (nevychýlenost, nezkreslenost)
vydatnost (eficience)
konzistence
dostatečnost

V praktických aplikacích, častěji než bodový odhad, určujeme intervalový odhad příslušného parametru. Tento odhad je reprezentován intervalem (T_D; T_H), v němž hledaný parametr leží s předem určenou pravděpodobností (spolehlivostí), kterou označujeme (1-α).

Intervaly spolehlivosti konstruujeme jako **jednostranné** nebo **dvoustranné**. V následující tabulce najdete přehled intervalových odhadů pro parametry normálního rozdělení včetně použitých výběrových charakteristik.

Intervaly spolehlivosti pro parametry normálního rozdělení

Odhadovaný parametr	**Vhodná výběrová charakteristika**	**Rozdělení výběrové char.**	**Meze oboustranného intervalu spolehlivosti**		**Dolní mez levostranného intervalu spolehlivosti**	**Horní mez pravostranného intervalu spolehlivosti**
			$\mathbf{T_D}$	$\mathbf{T_H}$	$\mathbf{T_D}$	$\mathbf{T_H}$
μ, známe σ	$Z=\frac{\overline{X}-\mu}{\sigma}\cdot\sqrt{n}$	N(0;1)	$\overline{X}-\frac{\sigma}{\sqrt{n}}\cdot z_{1-\frac{\alpha}{2}}$	$\overline{X}+\frac{\sigma}{\sqrt{n}}\cdot z_{1-\frac{\alpha}{2}}$	$\overline{X}-\frac{\sigma}{\sqrt{n}}\cdot z_{1-\alpha}$	$\overline{X}+\frac{\sigma}{\sqrt{n}}\cdot z_{1-\alpha}$
μ, neznáme σ	$T_{n-1}=\frac{\overline{X}-\mu}{S}\cdot\sqrt{n}$	t_{n-1}	$\overline{X}-\frac{S}{\sqrt{n}}\cdot t_{1-\frac{\alpha}{2},\,n-1}$	$\overline{X}+\frac{S}{\sqrt{n}}\cdot t_{1-\frac{\alpha}{2},\,n-1}$	$\overline{X}-\frac{S}{\sqrt{n}}\cdot t_{1-\alpha,\,n-1}$	$\overline{X}+\frac{S}{\sqrt{n}}\cdot t_{1-\alpha,\,n-1}$
σ^2	$\chi=\frac{(n-1)S^2}{\sigma^2}$	χ^2_{n-1}	$\frac{(n-1)}{x_{1-\frac{\alpha}{2},\,n-1}}\cdot S^2$	$\frac{(n-1)}{x_{\frac{\alpha}{2},\,n-1}}\cdot S^2$	$\frac{(n-1)}{x_{1-\alpha,\,n-1}}\cdot S^2$	$\frac{(n-1)}{x_{\alpha,\,n-1}}\cdot S^2$
σ	intervalový odhad je odvozen z intervalového odhadu σ^2		$\sqrt{\frac{(n-1)}{x_{1-\frac{\alpha}{2},\,n-1}}}\cdot S$	$\sqrt{\frac{(n-1)}{x_{\frac{\alpha}{2},\,n-1}}}\cdot S$	$\sqrt{\frac{(n-1)}{x_{1-\alpha,\,n-1}}}\cdot S$	$\sqrt{\frac{(n-1)}{x_{\alpha,\,n-1}}}\cdot S$
π	$P_1=\frac{p-\pi}{\sqrt{\pi(1-\pi)}}\cdot\sqrt{n}$	N(0;1)	$p-\sqrt{\frac{p\cdot(1-p)}{n}}\cdot z_{1-\frac{\alpha}{2}}$	$p+\sqrt{\frac{p\cdot(1-p)}{n}}\cdot z_{1-\frac{\alpha}{2}}$	$p-\sqrt{\frac{p\cdot(1-p)}{n}}\cdot$	$p+\sqrt{\frac{p\cdot(1-p)}{n}}\cdot z_{1-\alpha}$

Velikost výběru má přímý vliv na přesnost odhadu parametrů základního souboru: čím větší rozsah výběru, tím přesnější je intervalový odhad. Ekonomické a časové důvody nás však mnohdy nutí volit rozsah výběru co nejmenší. V praxi proto hledáme kompromis, který pro požadovanou přesnost výpočtu (**přípustnou chybu odhadu Δ)** povede k co nejmenšímu rozsahu výběru.

Odhadovaný parametr	Rozsah výběru
μ, známe σ	$n \geq \left(\frac{\sigma}{\Delta} \cdot z_{1-\frac{\alpha}{2}}\right)^2$
μ, neznáme σ	$n \geq \left(\frac{S_1}{\Delta} \cdot t_{1-\frac{\alpha}{2},\, n-1}\right)^2$
π	$n \geq \frac{p_1 \cdot (1-p_1)}{\Delta^2} \cdot z^2_{1-\frac{\alpha}{2}}$

Intervalové odhady můžeme použít také ke srovnávání středních hodnot, resp. relativních četností dvou populací:

Odhadovaný vztah mezi parametry	**Vhodná výběrová charakteristika**	**Rozdělení výběrové char.**	**Meze oboustranného intervalu spolehlivosti**		**Dolní mez levostranného intervalu spolehlivosti**	**Horní mez pravostranného intervalu spolehlivosti**
			T_D	**T_H**	**T_D**	**T_H**
μ_1- μ_1, známe σ_1; σ_2	$Z_2 = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	N(0;1)	$(\overline{X}_1 - \overline{X}_2) - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \cdot z_{1-\frac{\alpha}{2}}$	$(\overline{X}_1 - \overline{X}_2) + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \cdot z_{1-\frac{\alpha}{2}}$	$(\overline{X}_1 - \overline{X}_2) - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \cdot z_{1-\alpha}$	$(\overline{X}_1 - \overline{X}_2) + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \cdot z_{1-\alpha}$
μ_1- μ_1, neznáme σ_1; σ_2	$T_2 = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$t_{n_1+n_2-2}$	$(\overline{X}_1 - \overline{X}_2) - S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot t_{1-\alpha,\, n_1+n_2}$	$(\overline{X}_1 - \overline{X}_2) + S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot t_{1-\alpha,\, n_1+n_2-2}$	$(\overline{X}_1 - \overline{X}_2) - S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot t_{1-\alpha,\, n_1+n_2-2}$	$(\overline{X}_1 - \overline{X}_2) + S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot t_{1-\alpha,\, n_1+n_2-}$
π_1-π_2	$P_2 = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$	N(0;1	$(p_1 - p_2) - \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \cdot z_{1-\alpha}$	$(p_1 - p_2) + \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \cdot z_{1-\alpha}$	$(p_1 - p_2) - \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \cdot z$	$(p_1 - p_2) + \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \cdot z_{1-\frac{\alpha}{2}}$

10.1. Útvar kontroly podniku Edison testoval životnost žárovek. Kontroloři vybrali z produkce podniku náhodně 50 žárovek a došli k závěru, že průměrná doba života těchto 50-ti žárovek je 950 hodin a příslušná výběrová směrodatná odchylka doby života je 100 hodin. Určete 95%-ní interval spolehlivosti životnosti žárovek firmy Edison.

Řešení:

Chceme najít 95%-ní interval spolehlivosti pro střední hodnotu životnosti žárovek firmy Edison, přičemž neznáme směrodatnou odchylku životnosti těchto žárovek. Máme k dispozici informace pocházející z výběru o rozsahu 50 žárovek, tj. rozsah výběru je vyšší než 30 a proto k nalezení příslušného intervalového odhadu můžeme použít následující vztah (jde o intervalový odhad střední hodnoty pro známé σ, kde jsme položili σ=s) :

$$P\left(\overline{X} - \frac{S}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}} < \mu < \overline{X} + \frac{S}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

Spolehlivost intervalového odhadu: $1-\alpha = 0{,}95$

$\Rightarrow$ Hladina významnosti: $\alpha = 1 - 0{,}95 = 0{,}05$

$\Rightarrow$ $\frac{\alpha}{2} = 0{,}025; \quad 1 - \frac{\alpha}{2} = 0{,}975$

$\Rightarrow$ $z_{0{,}975} = 1{,}96$ (viz. Tabulka 1)

Výběrový soubor: $\overline{X} = 950\ hodin$

$S = 100\ hodin$

$n = 50$

Dosadíme: $$P\left(950 - \frac{100}{\sqrt{50}} \cdot 1{,}96 < \mu < 950 + \frac{100}{\sqrt{50}} \cdot 1{,}96\right) = 0{,}95$$

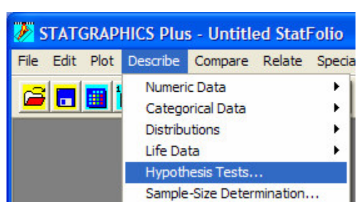
Po úpravě dostáváme: $P(922{,}3 < \mu < 977{,}7) = 0{,}95$

Tzn., že s 95%-ní spolehlivostí můžeme tvrdit, že životnost žárovek firmy Edison se pohybuje v rozmezí 922 hodin 18 minut až 977 hodin 42 minut.

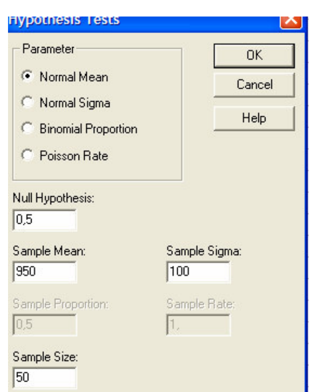
Řešení ve Statgraphicsu:

Statgraphics nám umožňuje určovat intervalové odhady těchto parametrů **normálního** rozdělení: střední hodnota, směrodatná odchylka, relativní četnost (podíl). Pro odhad střední hodnoty se používá pouze výběrová statistika T (tzn. že Statgraphics neumožňuje odhadovat střední hodnotu pro případ, kdy známe směrodatnou odchylku).

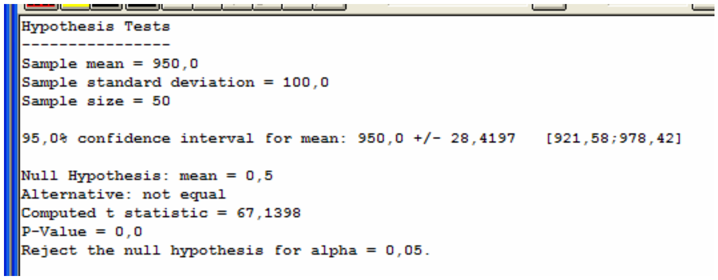
V našem případě chceme určit intervalový odhad střední hodnoty. V menu **Describe** zvolíme položku **Hypothesis Tests …**



V okně Hypothesi Test zadáme požadované údaje: **zaškrtneme pole Normal Mean** (střední hodnota normálního rozdělení), pole **Null Hypothesis** nás v tuto chvíli nezajímá, jako **Sample mean** (= výběrová střední hodnota = průměr) zadáme 950, jako **Sample Sigma** (= výběrová směrodatná odchylka) zadáme 100 a do pole **Sample Size** (= rozsah výběru) zapíšeme 50.



V textovém výstupu najdeme oboustranný 95% ní interval spolehlivosti pro střední hodnotu.



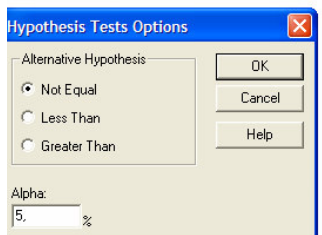
```
Hypothesis Tests
----------------
Sample mean = 950,0
Sample standard deviation = 100,0
Sample size = 50

95,0% confidence interval for mean: 950,0 +/- 28,4197   [921,58;978,42]

Null Hypothesis: mean = 0,5
Alternative: not equal
Computed t statistic = 67,1398
P-Value = 0,0
Reject the null hypothesis for alpha = 0,05.
```

Porovnejte zjištěné výsledky s „ručním“ výpočtem. (Rozdíly jsou způsobeny přesnějším stanovením příslušných kvantilů).

Chceme-li nastavit jinou spolehlivost odhadu (resp. hladinu významnosti), popřípadě chceme.li určit některý z jednostranných odhadů, nastavení provedeme v okně **Analysis Option** (RC na textový výstup).



Je zřejmé, že v poli **Alpha** nastavíme hladinu významnosti. Je-li zaškrtnuto pole **Not Equal** (není rovno), Statgraphics poskytne jako výstup oboustranný interval spolehlivosti, zaškrtneme-li **Less Than** (menší než), dostaneme pravostranný interval spolehlivosti, resp. jeho horní mez a obdobně zaškrtneme-li **Greather Than** (větší než), dostaneme levostranný interval spolehlivosti, resp. jeho dolní mez.

10.2. Obchodní řetězec TETO si v dubnu 2006 zadal studii týkající se počtu zákazníku v prodejně TETO Poruba v pátek odpoledne (od 12:00 do 18:00) hodin. Po jednom měsíci sledování prodejny jsme získali tyto údaje:

Datum	**Počet zákazníků v TETO Poruba (12:00-18:00) hodin**
2.5.2006	**3756**
9.5.2006	**2987**
16.5.2006	**3042**
23.5.2006	**4206**
30.5.2006	**3597**

a) Objasněte, proč jsme nezískali výběrový soubor o rozsahu alespoň 30 hodnot a jaké jsou důsledky volby výběru o malém rozsahu.
b) Určete pro managment řetězce TETO 95%-ní interval spolehlivosti počtu zákazníku v prodejně TETO Poruba v pátek odpoledne.

Řešení:

ada) Pro získání výběru o rozsahu minimálně 30 hodnot bychom museli danou prodejnu sledovat minimálně 30 pátku (tj. déle než půl roku), což by vedlo jednak k zvýšení finanční náročnosti studie, jednak bychom museli dlouho čekat na výsledky. Z těchto důvodu jsme zvolili menší rozsah výběru (n=5) odpovídající měsíčnímu sledování prodejny. Nevýhodou malého rozsahu výběru je nízká přesnost odhadu (poměrně široký interval).

adb) Určujeme intervalový odhad střední hodnoty s neznámou směrodatnou odchylkou a malým rozsahem výběru, proto pro jeho výpočet použijeme následující vztah:

$$P\left(\overline{X} - \frac{s}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2},\, n-1} < \mu < \overline{X} + \frac{s}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2},\, n-1}\right) = 1-\alpha$$

Spolehlivost intervalového odhadu: $1-\alpha = 0{,}95$

$\Rightarrow$ Hladina významnosti: $\alpha = 1 - 0{,}95 = 0{,}05$

$\Rightarrow$ $\frac{\alpha}{2} = 0{,}025; \quad 1 - \frac{\alpha}{2} = 0{,}975$

$\Rightarrow$ $t_{0{,}975,\, 4} = 2{,}78$ (viz. Tabulka 2)

Výběrový soubor:

$$\bar{x} = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{3756 + 2987 + 3042 + 4206 + 3597}{5} = 3517{,}6$$

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{(3756 - 3517{,}6)^2 + \ldots + (3597 - 3517{,}6)^2}{4} = 261191{,}3 \quad \Rightarrow \quad s = 511{,}1$$

$n = 5$

Dosadíme: $$P\left(3517{,}6 - \frac{511{,}1}{\sqrt{5}} \cdot 2{,}78 < \mu < 3517{,}6 + \frac{511{,}1}{\sqrt{5}} \cdot 2{,}78\right) = 0{,}95$$

Po úpravě dostáváme: $P(2882{,}2 < \mu < 4153{,}0) = 0{,}95$

Tzn., že s 95%-ní spolehlivostí můžeme tvrdit, že návštěvnost TETO Poruba se v libovolný pátek v odpoledních hodinách bude pohybovat v rozmezí 2882 až 4153 zákazníků.

Část b vyřešte pomocí Statgraphicsu (viz. př. 10.1.)

10.3. Automat vyrábí pístové kroužky o daném průměru. Při kontrole kvality bylo náhodně vybráno 80 kroužků a vypočtena směrodatná odchylka jejich průměru 0,04mm. Odhadněte 95%-ní levostranný interval spolehlivosti pro rozptyl a směrodatnou odchylku průměru pístových kroužků.

Řešení:

Nejdříve najdeme 95%-ní levostranný interval spolehlivosti pro rozptyl. Pro jeho nalezení použije následující vztah:

$$P\left(\frac{(n-1)}{x_{1-\alpha,\, n-1}}\cdot S^2 < \sigma^2\right) = 1-\alpha$$

Spolehlivost intervalového odhadu: $1-\alpha = 0{,}95$

$\Rightarrow \qquad x_{0{,}95;\,79} \cong 100{,}7 \qquad$ (viz. Tabulka 3)

Výběrový soubor: $S^2 = (0{,}04)^2 = 0{,}0016\ mm^2$

n = 80

Po dosazení: $P\left(\frac{79}{100{,}7}\cdot 0{,}0016 < \sigma^2\right) = 0{,}95$

$$P(0{,}0013 < \sigma^2) = 0{,}95$$

Jednoduchou úpravou pak získáme 95%-ní levostranný interval spolehlivosti pro směrodatnou odchylku:

$$P\left(\sqrt{0{,}0013} < \sigma\right) = 0{,}95$$

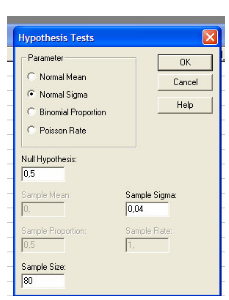
$$P(0{,}035 < \sigma) = 0{,}95$$

S 95%-ní spolehlivostí tedy můžeme tvrdit, že rozptyl průměru pístových kroužků je větší než $2{,}2.10^{-3}$ mm^2 (resp., že s 95%-ní spolehlivostí je směrodatná odchylka průměru pístových kroužků větší než $4{,}6.10^{-2}$ mm).

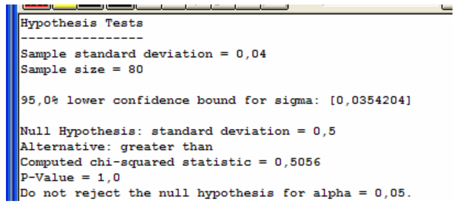
Řešení ve Statgraphicsu:

Statgraphics nám umožňuje nalézt intervalové odhady pro směrodatnou odchylku, nikoliv pro rozptyl.

Zvolíme **menu Describe**, položku **Hypothesis Tests …** . V **okně Hypothesis Tests** zaškrtneme **Normal sigma** a doplníme příslušné parametry výběru. Pole **Null Hypothesis** nás opět nezajímá, v poli **Sample Sigma** (výběrová směrodatná odchylka) doplníme hodnotu 0,04 a jako **Sample Size** (rozsah výběru) dosadíme 80.



Automaticky se nám vygeneroval oboustranný 95% ní interval spolehlivosti pro sigma, my však požadujeme interval levostranný, tj. pouze dolní mez tohoto intervalu (uvedená mez má být s požadovanou spolehlivostí menší než skutečná směrodatná odchylka). Proto v okně **Analysis Option** (RC na textový výstup) zaškrtneme **Greather than** (větší než).



```
Hypothesis Tests
----------------
Sample standard deviation = 0,04
Sample size = 80

95,0% lower confidence bound for sigma: [0,0354204]

Null Hypothesis: standard deviation = 0,5
Alternative: greater than
Computed chi-squared statistic = 0,5056
P-Value = 1,0
Do not reject the null hypothesis for alpha = 0,05.
```

Je zřejmé, že: $P(0{,}035 < \sigma) = 0{,}95$. Chtěli-li bychom požadovaný interval spolehlivosti určit pro rozptyl, musíme udanou dolní mez umocnit na druhou.

10.4. Při kontrole data spotřeby určitého druhu masové konzervy ve skladech produktů masného průmyslu bylo náhodně vybráno 320 konzerv a zjištěno, že 59 z nich má prošlou záruční lhůtu. Stanovte 95% interval spolehlivosti pro odhad procenta konzerv s prošlou záruční lhůtou.

Řešení:

Pro nalezení 95%-ního intervalu spolehlivosti pro relativní četnost použijeme následující vztah:

$$P\left(p - \sqrt{\frac{p\cdot(1-p)}{n}}\cdot z_{1-\frac{\alpha}{2}} < \pi < p + \sqrt{\frac{p\cdot(1-p)}{n}}\cdot z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

Spolehlivost intervalového odhadu: $1-\alpha = 0{,}95$

$\Rightarrow$ Hladina významnosti: $\alpha = 1-0{,}95 = 0{,}05$

$\Rightarrow$ $\frac{\alpha}{2} = 0{,}025; \quad 1-\frac{\alpha}{2} = 0{,}975$

$\Rightarrow$ $z_{0{,}975} = 1{,}96$ (viz. Tabulka 1)

Výběrový soubor: $p = \frac{59}{320} \cong 0{,}18$

$n = 320$

Po dosazení:

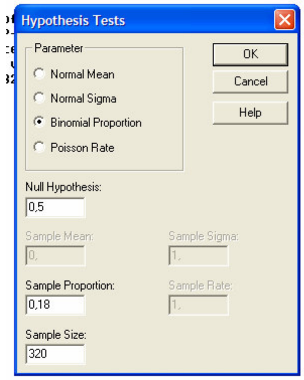
$$P\left(0{,}18 - \sqrt{\frac{0{,}18\cdot(1-0{,}18)}{320}}\cdot 1{,}96 < \pi < 0{,}18 + \sqrt{\frac{0{,}18\cdot(1-0{,}18)}{320}}\cdot 1{,}96\right) = 0{,}95$$

$$P(0{,}138 < \pi < 0{,}222) = 0{,}95$$

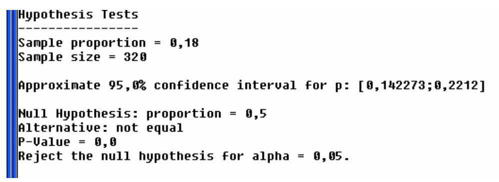
S 95%-ní spolehlivostí můžeme tvrdit, že mezi masovými konzervami se v daném skladu nachází mezi 13,8% a 22,2% konzerv s prošlou záruční lhůtou.

Řešení ve Statgraphicsu:

Opět zvolíme **menu Describe**, položku **Hypothesis Tests …** . V **okně Hypothesis Tests** zaškrtneme **Binomial Proportion** a doplníme příslušné parametry výběru. Pole **Null Hypothesis** nás opět nezajímá, v poli **Sample Proportion** (výběrová relativní četnost, výběrový podíl) doplníme hodnotu 0,18 a jako **Sample Size** (rozsah výběru) dosadíme 320.



Automaticky se nám vygeneroval oboustranný 95% ní interval spolehlivosti pro relativní četnost:



```
Hypothesis Tests
----------------
Sample proportion = 0,18
Sample size = 320

Approximate 95,0% confidence interval for p: [0,142273;0,2212]

Null Hypothesis: proportion = 0,5
Alternative: not equal
P-Value = 0,0
Reject the null hypothesis for alpha = 0,05.
```

Rozdíly oproti vypočteným hodnotám jsou způsobeny zaokrouhlovaním.

10.5. Výběrovým šetřením bychom chtěli odhadnout průměrnou mzdu pracovníků určitého výrobního odvětví. Z vyčerpávajícího šetření, které probíhalo před několika měsíci, víme, že směrodatná odchylka mezd byla 750,-Kč. Odhad chceme provést s 95% spolehlivosti a jsme ochotni připustit maximální chybu ve výši 50,-Kč. Jak velký musíme provést výběr, abychom zajistili požadovanou přesnost a spolehlivost?

Řešení:

Chceme odhadnou rozsah výběru pro intervalový odhad střední hodnoty známe-li směrodatnou odchylku σ (vyčerpávající šetření = zkoumání celého základního souboru (populace)).

$$1-\alpha = 0{,}95 \Rightarrow \alpha = 0{,}05 \Rightarrow 1-\frac{\alpha}{2} = 0{,}975$$

$$z_{0,975} = 1{,}96 \quad (\mathit{Tabulka}\ 1)$$

$$\sigma = 750\ K\check{c}$$

$$\Delta \leq 50\ K\check{c}$$

Rozsah výběru odhadneme v tomto případě podle vztahu:

$$n \geq \left(\frac{\sigma}{\Delta} \cdot z_{1-\frac{\alpha}{2}} \right)^2$$

Po dosazení:

$$n \geq \left(\frac{750}{50} \cdot 1{,}96 \right)^2$$
$$n \geq 864{,}4$$

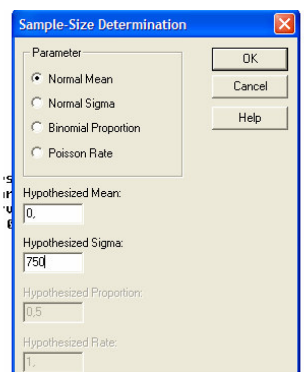
Chceme-li dosáhnout přípustné chyby ve výši maximálně 50,- Kč, musíme pro nalezení 95%-ního intervalového odhadu provést výběrové šetření na souboru o rozsahu minimálně 865 pracovníků.

Řešení ve Statgraphicsu:

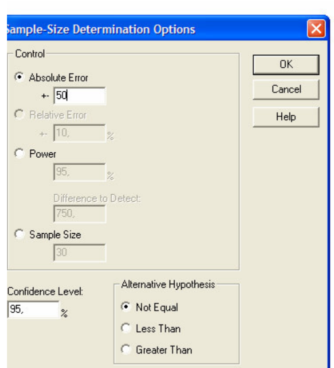
Statgraphics nám umožňuje odhadnout rozsah výběru na základě požadované maximální přípustné chyby (resp. relativní chyby). A to pro odhad střední hodnoty, směrodatné odchylky a relativní četnosti.

Zvolíme **menu Describe**, **položku Sample – Size Determination …** .

V okně **Sample – Size Determination …** zaškrtneme Normal Mean a nastavíme zadané parametry (tj. **Hypothesis Sigma** (odhadovaná směrodatná odchylka) – 750).



V okně **Sample – Size Determination Options** zadáme velikost přípustné chyby v poli, které se zvýrazní po označení **Absolute Error** (= 50). Spolehlivost odhadu (Confidence Level) se shoduje s přednastavenou hodnotou (95%).



V textovém výstupu pak najdeme hledanou informaci:

```
Sample-Size Determination
-------------------------

Parameter to be estimated: normal mean
Desired tolerance: +- 50,0
Confidence level: 95,0%
Assumed sigma: 750,0

The required sample size is n=865 observations.


The StatAdvisor
---------------
   This procedure determines the sample size required when estimating
the mean of a normal distribution.  Assuming that the standard
deviation of the normal distribution equals 750,0, 865 observations
are required to estimate the mean to within +/-50,0 with 95,0%
confidence.
```

10.6. Diskety dvou velkých výrobců - Sonik a 5M byly podrobeny zkoušce kvality. Diskety obou výrobců jsou baleny po 20-ti kusech. Ve 40-ti balíčcích fy Sonik bylo nalezeno 24 vadných disket, ve 30-ti balíčcích 5M bylo nalezeno 14 vadných disket. Určete 95%-ní interval spolehlivosti pro rozdíl v procentu vadných disket v celkové produkci firem Sonik a 5M.

Řešení:

Označme si procento vadných disket v produkci fy Sonik π_1 a procento vadných disket v produkci fy 5M π_2.

Pro určení požadovaného intervalu použijeme vztah:

$$P\left((p_1 - p_2) - \sqrt{p(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \cdot z_{1-\frac{\alpha}{2}} < (\pi_1 - \pi_2) < (p_1 - p_2) + \sqrt{p(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \cdot z_{1-\frac{\alpha}{2}}\right) =$$

$$= 1-\alpha$$

Spolehlivost intervalového odhadu: $1-\alpha = 0{,}95$

$\Rightarrow$ Hladina významnosti: $\alpha = 1 - 0{,}95 = 0{,}05$

$\Rightarrow$ $\frac{\alpha}{2} = 0{,}025; \quad 1-\frac{\alpha}{2} = 0{,}975$

$\Rightarrow$ $z_{0,975} = 1{,}96$ (viz. Tabulka 1)

Výběrové soubory:

Sonik:

$$x_1 = 24$$

$$n_1 = 40 \cdot 20 = 800$$

$$p_1 = \frac{24}{800} = 0{,}030 \quad (\textit{výběrový podíl vadných disket fy Sonik})$$

5M:

$$x_2 = 14$$
$$n_1 = 30 \cdot 20 = 600$$
$$p_1 = \frac{14}{600} = 0{,}023 \quad (\textit{výběrový podíl vadných disket fy 5M})$$

$$p = \frac{24+14}{800+600} = 0{,}027$$

Po dosazení:

$$P((0{,}007) - 0{,}017 < (\pi_1 - \pi_2) < (0{,}007) + 0{,}017) = 0{,}95$$

$$P(-0{,}010 < (\pi_1 - \pi_2) < 0{,}024) = 0{,}95$$

$$P(-1{,}0\ \% < (\pi_1 - \pi_2) < 2{,}4\ \%) = 0{,}95$$

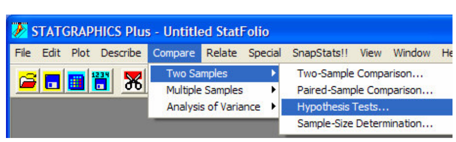
S 95%-ní spolehlivostí můžeme tvrdit, že rozdíl mezi podílem vadných disket firmy Sonik a podílem vadných disket firmy 5M je v rozmezí –1,0 % a 2,4%. Tzn., že nemůžeme říci, které diskety jsou kvalitnější.

V případě, že by rozdíl mezi podílem vadných disket firmy Sonik a podílem vadných disket firmy 5M byl záporný $(\pi_1 - \pi_2 < 0)$, znamenalo by to, že diskety firmy Sonik jsou kvalitnější (obsahují menší podíl vadných) než diskety firmy 5M $(\pi_1 < \pi_2)$. Obdobně v případě, že by rozdíl mezi podílem vadných disket firmy Sonik a podílem vadných disket firmy 5M byl kladný $(\pi_1 - \pi_2 > 0)$, znamenalo by to, že diskety firmy Sonik mají horší kvalitu (obsahují větší podíl vadných) než diskety firmy 5M $(\pi_1 > \pi_2)$. V našem případě víme, že rozdíl mezi podílem vadných disket firmy Sonik a podílem vadných disket firmy 5M může být jak kladný, tak i záporný a proto nemůžeme říci, které diskety jsou kvalitnější. Ale to už jsme se dostali k testování hypotéz, jimž se budeme zabývat v následující kapitole.

Řešení ve Statgraphicsu:

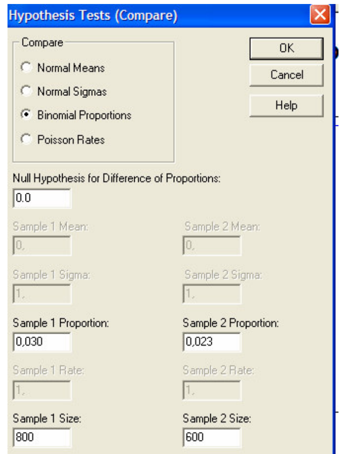
Také porovnávání parametrů dvou normálních rozdělení umožňuje Statgraphics. Můžeme zde hledat intervalové odhady pro rozdíl středních hodnot dvou populací, pro rozdíl podílů dvou populací a pro poměr mezi směrodatnými odchylkami dvou populací.

Výše uvedené intervalové odhady najdeme v **menu Compare**, v položce **Hypothesis Tests …**



V okně **Hypothesis Tests (Compare)** zaškrtneme pole **Binomial Proportion** a vyplníme požadované parametry – **Sample 1 Proportion** (výběrový podíl pro 1. výběr (Sonik) = 0,030), **Sample 1 Size** (rozsah výběru pro 1. výběr (Sonik) = 800), **Sample 2 Proportion** (výběrový podíl pro 2. výběr (5M) = 0,023), **Sample 2 Size** (rozsah výběru pro 2. výběr (5M)

= 600). Hodnota uvedená v poli **Null Hypothesis for Diference of Proportions** nemá na určení intervalového odhadu pro rozdíl podílů význam.



V textovém výstupu najdeme oboustranný interval spolehlivosti pro rozdíl podílů. Chtěli-li bychom najít jednostranné odhady, popř. změnit hladinu významnosti (spolehlivost odhadu), změnu provedeme v okně **Analysis Option** (RC na textový výstup).