

4.1 Bodový odhad

Princíp bodového odhadu spočíva v odhade neznámych parametrov (strednej hodnoty, rozptylu, smerodajnej odchýlky, atď.) prostredníctvom výberových charakteristík, ktoré sú reprezentované jedným číslom (bodom).

POJMY

- | | |
|-------------------------------|--------------------------------|
| ✓ Základný súbor | ✓ Výberový priemer |
| ✓ Výberový súbor | ✓ Výberový rozptyl |
| ✓ Výberové skúmanie | ✓ Výberová smerodajná odchýlka |
| ✓ Parametre základného súboru | ✓ Výberový podiel |
| ✓ Výberové charakteristiky | ✓ Vlastnosti bodového odhadu |
| ✓ Estimátor | |

VZORCE

Odhad strednej hodnoty pomocou výberového priemeru

$$\text{est } \mu = \bar{x}$$

$$\bar{x} = \frac{\sum_{j=1}^n x_j}{n}$$

kde: x_i – hodnota štatistického znaku

$\bar{x}$ – aritmetický priemer

n_i – absolútna početnosť

m – počet tried

Odhad rozptylu pomocou výberového rozptylu

$$\text{est } \sigma^2 = s_1^2$$

$$s_1^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

kde: x_i – hodnota štatistického znaku

$\bar{x}$ – aritmetický priemer

n_i – absolútna početnosť

m – počet tried

Odhad smerodajnej odchýlky pomocou výberovej smerodajnej odchýlky

$$\text{est } \sigma = s_1$$

$$s_1 = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

kde: x_j – hodnota štatistického znaku

$\bar{x}$ – aritmetický priemer

Odhad podielu pomocou výberového podielu

$$\text{est } \pi = p$$

$$p = \frac{n_1}{n}$$

kde: x_j – hodnota štatistického znaku

$\bar{x}$ – aritmetický priemer

FUNKCIE V EXCELI

= **RANDBETWEEN** (dolná hranica, horná hranica)
= **AVERAGE** (číslo1, číslo2,...)
= **VAR** (číslo1, číslo2,...)
= **STDEV** (číslo1, číslo2,...)

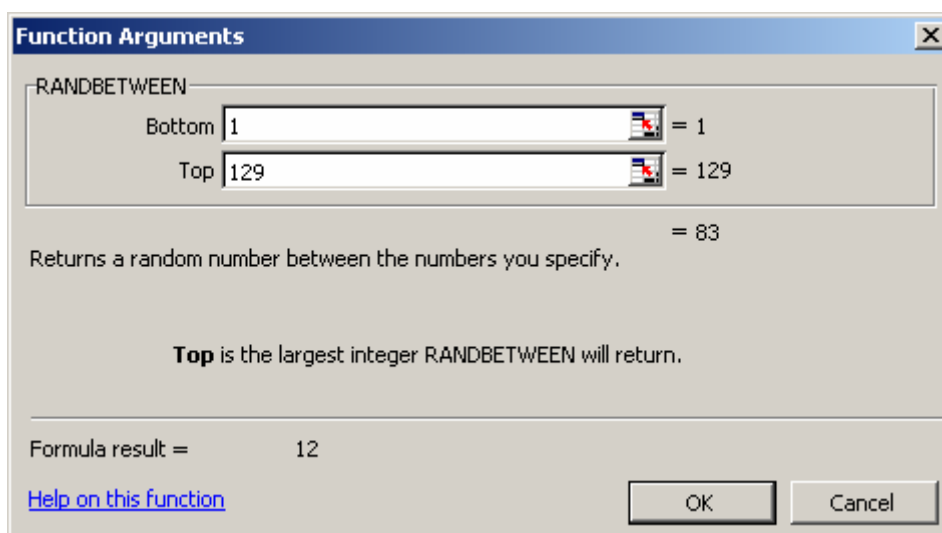
RIEŠENÝ PRÍKLAD

Úloha: Uskutočniť výberové skúmanie, pričom základným súborom ostáva súbor 129 doteraz analyzovaných poľnohospodárskych podnikov. Z výberového súboru následne vypočítať výberové charakteristiky, ktoré budú bodovým odhadom parametrov celého základného súboru.

Zadanie: Vytvorte náhodným spôsobom výberový súbor s rozsahom 20 štatistických jednotiek, z ktorého vypočítajte výberové charakteristiky. Analýzy realizujte na štatistickom znaku priemerný mesačný zárobok. Na základe výberového súboru urobte odhad podielu družstiev v skúmanom súbore.

Riešenie: Prvým zadaniem je vytvoriť výberový súbor. Budeme vychádzať z jednoduchého náhodného výberu, tzn., že zo základného súboru 129 poľnohospodárskych podnikov náhodne vyberieme 20 podnikov, ktoré budú tvoriť výberový súbor. Pri vytváraní je možné využiť matematickú funkciu RANDBETWEEN (slide 4.1), ktorá po zadaní minimálnej (1) a maximálnej (129) hodnoty náhodne nagenereuje číslo. Po skopírovaní tejto hodnoty do ďalších buniek (n=20, t.j. musíme nakopírovať ešte 19 ďalších hodnôt) dostávame jednotky¹ výberového súboru.

Slide 4.1



¹ Nagenеровané hodnoty vo výberovom súbore predstavujú čísla podnikov, u ktorých budeme skúmať priemerné mesačné zárobky.

Čísla, ktoré sme získali, predstavujú poradové čísla podnikov, na základe ktorých vyberieme podniky do výberového súboru. Vytvorený výberový súbor vyzerá potom nasledovne:

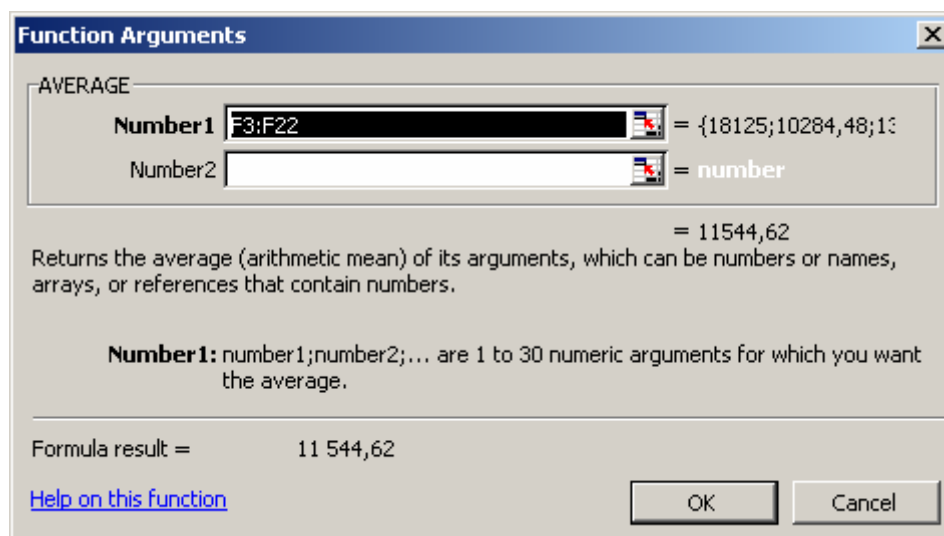
Tabuľka 1: Výberový súbor priemerných mesačných zárobkov za rok 2005

Porad.č.	Číslo podniku	Forma	ROK2005
83	4965	ostatné	18125
40	4018	družstvo	10284,48
35	3634	družstvo	13250
20	3624	družstvo	14797,22
79	4035	ostatné	13583,33
37	4017	ostatné	16181,37
129	5012	ostatné	7125
85	4966	družstvo	15098,48
34	4016	ostatné	10622,39
113	4995	ostatné	14988,09
86	4967	ostatné	8811,9
115	4996	družstvo	5020,83
87	4041	ostatné	11242,06
99	4981	družstvo	11107,25
25	3627	družstvo	4797,61
126	4808	ostatné	9535,71
30	3630	ostatné	11942,3
103	4793	družstvo	8758,62
71	4031	ostatné	12966,66
8	3614	ostatné	12654,1

Po vytvorení súboru môžeme prísť k výpočtu výberových charakteristík. Vychádzame z malého rozsahu hodnôt (20), t.j. údaje nie je potrebné triediť. Pri výpočte využijeme funkcie, ktoré nám Excel ponúka.²

Výberový priemer

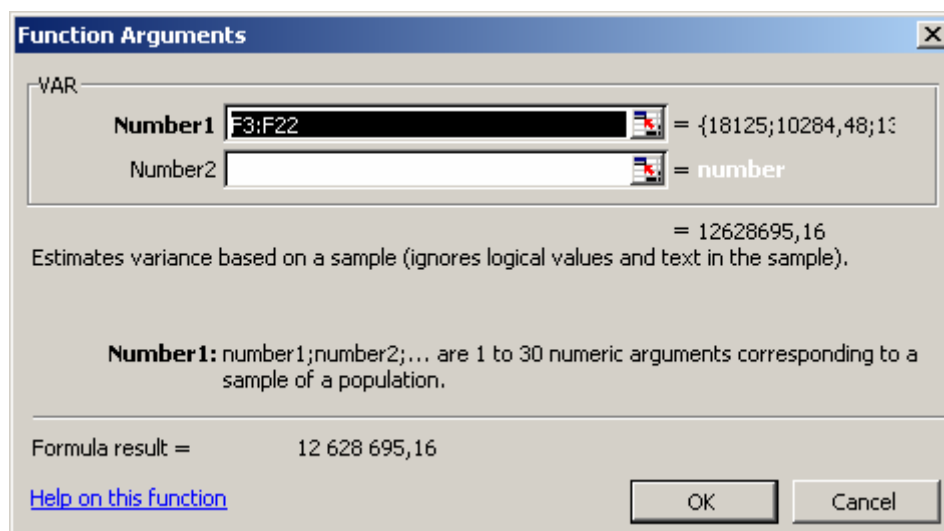
Použijeme funkciu AVERAGE. Hodnota výberového priemeru sa rovná 11 544,62. Interpretácie uvedieme opäť v závere po dopočítaní ostatných výberových charakteristík.



Výberový rozptyl

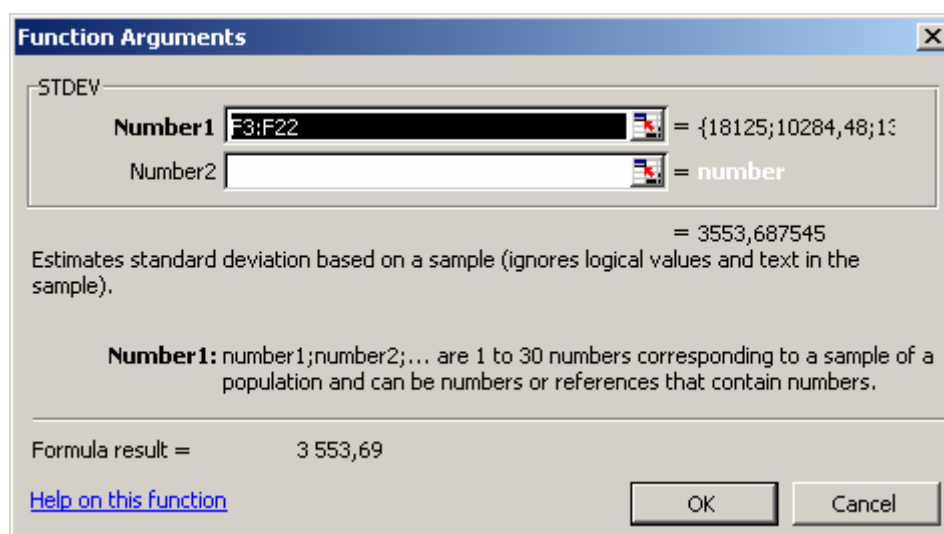
² Pri výpočte by bolo možné vypočítať charakteristiky prostredníctvom vzorcov alebo využiť Data Analysis/Summary Statistics (kapitola 2).

Na výpočet tejto výberovej charakteristiky slúži funkcia VAR. Výsledná hodnota je 12 628 695,16.



Výberová smerodajná odchýlka

Smerodajnú odchýlku by bolo možné dopočítať prostredníctvom dvoch funkcií. Využitím funkcie SQRT, t.j. odmocnili by sme výberový rozptyl alebo pomocou funkcie STDEV. Výberová smerodajná odchýlka sa rovná 3 553,69.



Výberový podiel

Poslednou úlohou v rámci bodových odhadov bolo na základe výberového súboru odhadnúť podiel družstiev v celom základnom súbore, ktorý určíme na základe relatívnej početnosti. V súbore máme 20 (n) podnikov, z nich je 8 (n1) družstiev. Bodový odhad vypočítame ako $8/20=0,4$, t.j. 40%.

Interpretácia výsledkov:

Pri interpretácii vychádzame z výsledkov, ktoré sme vypočítali z výberového súboru s cieľom robiť závery na celý základný súbor. Čiže na základe výberového priemeru odhadujeme, že priemerné mesačné zárobky vo všetkých poľnohospodárskych podnikoch (základný súbor) v roku 2005 boli 11 544,62 Sk. Odhaduje, že variabilita v základnom súbore meraná

rozptylom bola 12 628 695,16 Sk², t.j. priemerné mesačné zárobky pracovníkov poľnohospodárskych podnikov kolísali v roku 2005 $\pm$ 3 553,69 Sk od priemeru. Vychádzajúc z výberového podielu odhadujeme, že z celkového počtu 129 poľnohospodárskych podnikov tvorili v roku 2005 družstvá 40% podnikov.

Výhodou bodového odhadu je, že výsledný odhad neznámeho parametra je tvorený jednou hodnotou. Keďže však bol výberový súbor vytváraný náhodným spôsobom, tak aj výberové charakteristiky sú vlastne náhodné premenné. Keby sme z toho istého základného súboru vytvorili ďalší výberový súbor náhodným spôsobom, je vysoko pravdepodobné, že v novom výberovom súbore by sa nachádzali iné štatistické jednotky, t.j. pri odhade by sme sa dopracovali k iným hodnotám výberových charakteristík. Z uvedeného vyplýva, že bodový odhad je síce presný (výsledkom je jedno číslo), ale málo spoľahlivý. Z tohto dôvodu sa častejšie využívajú intervalové odhady.

4.2 Intervalový odhad

Ako už bolo vyššie spomenuté výhodou intervalových odhadov je ich spoľahlivosť. Podstata intervalového odhadu spočíva v určení hraníc intervalu spoľahlivosti pri zadanej pravdepodobnosti.

POJMY	
✓ Parametre základného súboru	✓ Interval spoľahlivosti
✓ Výberové charakteristiky	✓ Prípustná chyba
✓ Spoľahlivosť odhadu	✓ Normované normálne rozdelenie
✓ Koeficient spoľahlivosti	✓ Studentovo t- rozdelenie
✓ Riziko odhadu	✓ chí kvadrát rozdelenie
VZORCE	

Interval spoľahlivosti pre strednú hodnotu: $n > 30$

$$P(\bar{x} - \Delta \leq \mu \leq \bar{x} + \Delta) = 1 - \alpha$$

$$\Delta = u_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

kde: $\bar{x}$ - výberový priemer
 Δ - prípustná chyba
 $u_{1-\alpha/2}$ - kvantil normovaného normálneho rozdelenia
 σ - smerodajná odchýlka

Interval spoľahlivosti pre strednú hodnotu: $n \leq 30$

$$P(\bar{x} - \Delta \leq \mu \leq \bar{x} + \Delta) = 1 - \alpha$$

$$\Delta = t_{\alpha} \cdot \frac{s_1}{\sqrt{n}}$$

kde: $\bar{x}$ - výberový priemer
 Δ - prípustná chyba
 t_{α} - kvantil Studentovho t-rozdelenia
 s_1 - výberová smerodajná

Interval spoľahlivosti pre rozptyl

$$P\left(\frac{(n-1) \cdot s_1^2}{\chi_{\alpha/2}^2} \leq \sigma \leq \frac{(n-1) \cdot s_1^2}{\chi_{1-\alpha/2}^2}\right) = 1 - \alpha$$

kde: $\chi_{1-\alpha/2}^2$ - kvantil chí kvadrát rozdelenia
 $\chi_{\alpha/2}^2$ - kvantil chí kvadrát rozdelenia
 s_1^2 - výberový rozptyl

Interval spoľahlivosti pre smerodajnú odchýlku

$$P\left(\sqrt{\frac{(n-1) \cdot s_1^2}{\chi_{\alpha/2}^2}} \leq \sigma \leq \sqrt{\frac{(n-1) \cdot s_1^2}{\chi_{1-\alpha/2}^2}}\right) = 1 - \alpha$$

kde: $\chi_{1-\alpha/2}^2$ - kvantil chí kvadrát rozdelenia
 $\chi_{\alpha/2}^2$ - kvantil chí kvadrát rozdelenia
 s_1^2 - výberový rozptyl

Interval spoľahlivosti pre podiel

Pre veľké výbery ak platí vzťah: $n > \frac{9}{p \cdot (1-p)}$

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

kde: p - výberový podiel
 $u_{1-\alpha/2}$ - kvantil normovaného normálneho rozdelenia
 π - odhadovaný podiel ZS
 n - rozsah výberového súboru

Pre malé výbery ak neplatí vzťah: $n > \frac{9}{p \cdot (1-p)}$

Využitím Fischerovho F rozdelenia

$$n_2 p_d / [(n_1 + 1) \cdot (1 - p_d)] = F_{(1-\alpha/2, 2(n_1+1), 2n_2)}$$

kde: n - rozsah výberového súboru
 n_1 - početnosť výskytu skúmaného znaku
 $n_2 = n - n_1$
 p_d - dolná hranica intervalu spoľahlivosti
 p_h - horná hranica intervalu spoľahlivosti

$F_{(1-\alpha/2, 2(n_1+1), 2n_2)}$ - kvantil Fischerovho rozdelenia s príslušnými stupňami

$$n_2 p_h / [(n_1 + 1) \cdot (1 - p_h)] = F_{(\alpha/2, 2(n_1+1), 2n_2)}$$

Rozsah výberového súboru

$$n = u_{1-\alpha/2}^2 \frac{\sigma^2}{\Delta^2}$$

$$n = u_{1-\alpha/2}^2 \frac{\pi(1-\pi)}{\Delta^2}$$

kde: n – rozsah výberového súboru
 $u_{1-\alpha/2}$ – kvantil normovaného normálneho rozdelenia
 π – odhad podielu pomocou p

FUNKCIE V EXCELI

= AVERAGE (číslo1, číslo2,...)
 = VAR (číslo1, číslo2,...)
 = STDEV (číslo1, číslo2,...)
 = CONFIDENCE (alfa, smerodajná odchýlka, rozsah výberového súboru)
 = NORMSINV (pravdepodobnosť)
 = TINV (pravdepodobnosť, stupne voľnosti)
 = CHINV (pravdepodobnosť, stupne voľnosti)
 = FIINV (pravdepodobnosť, stupne voľnosti1, stupne voľnosti2)

RIEŠENÝ PRÍKLAD

Úloha: Vychádzajúc z výberových charakteristík urobiť intervalový odhad, na základe ktorého odhadneme neznáme parametrov celého základného súboru pri zadanej spoľahlivosti.

Zadanie: Vypočítajte:

- 95% a 99% interval spoľahlivosti pre strednú hodnotu, rozptyl a smerodajnú odchýlku priemerných mesačných zárobkov.
- Určite v akých hraniciach sa bude pohybovať podiel družstiev z celkového počtu podnikov pri 95% a 99% pravdepodobnosti.

Riešenie:

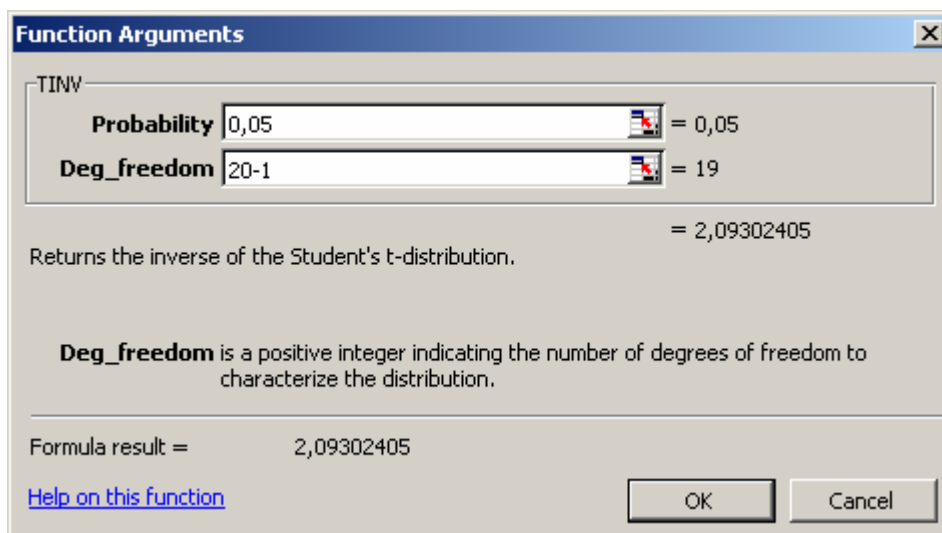
Pri určovaní intervalu spoľahlivosti je potrebné vypočítať hranice intervalu, v ktorých sa bude neznámy parameter pohybovať s vopred stanovenou pravdepodobnosťou. V našom príklade chceme odhadnúť hranice intervalov s 95%, resp. 99% spoľahlivosťou.

Interval spoľahlivosti pre strednú hodnotu

Interval spoľahlivosti pre strednú hodnotu sa určí, keď k bodovému odhadu strednej hodnoty (výberovému priemeru) pričítame (výpočet hornej hranice), resp. odčítame (výpočet dolnej hranice) prípustnú chybu (delta). Prípustná chyba teda predstavuje polovicu šírky intervalu spoľahlivosti a jej veľkosť je závislá na zvolenej spoľahlivosti ($u_{1-\alpha/2}$, t_α), variabilite základného súboru (σ) a rozsahu výberového súboru (n). Opäť je možné vypočítať prípustnú chybu buď mechanicky, prostredníctvom vzorca alebo pomocou funkcie.

a) výpočet pomocou vzorca

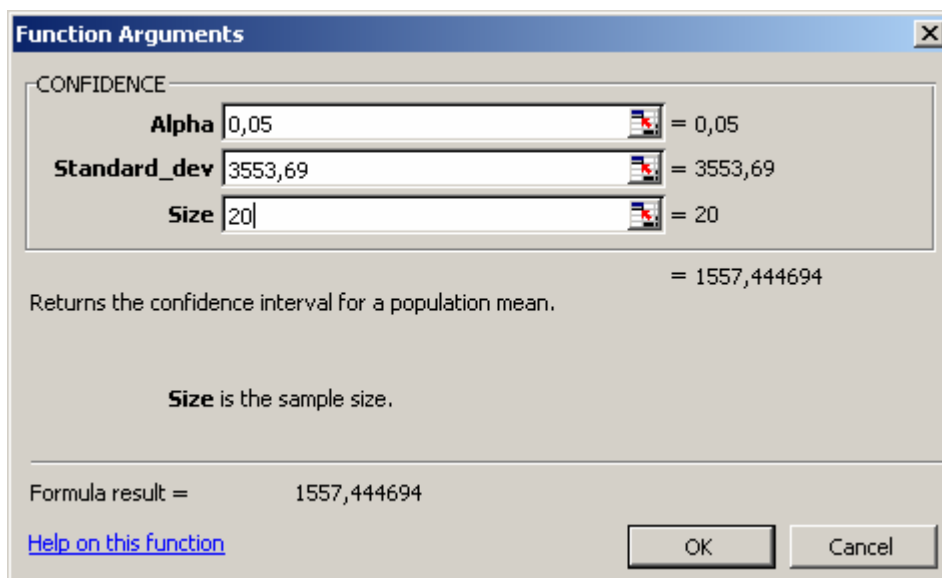
Do vzorca pre výpočet prípustnej chyby potrebuje poznať kritickú hodnotu známeho teoretického rozdelenia, ktorú určíme prostredníctvom funkcie. Keďže rozsah výberového súboru je menší ako 30 ($n=20$), budeme pri výpočte vychádzať zo Studentovho t-rozdelenia (funkcie TINV). Prvý parameter funkcie predstavuje zvolenú pravdepodobnosť (v prípade funkcie TINV sa pravdepodobnosť rovná α^3) a druhý parameter je parametrom Studentovho t-rozdelenia, ktorý sa určí ako $n-1$.



b) výpočet pomocou funkcie

Na výpočet prípustnej chyby slúži v Exceli funkcia CONFIDENCE. Túto funkcie je možné použiť len vtedy, ak je známa variabilita základného súboru, resp. ak je $n \geq 30$, t.j. ak vychádzame z normovaného normálneho rozdelenia. Vychádzajúc z údajov v príklade by použitie funkcie vyzeralo nasledovne:

³ Využitie funkcie prezentujeme pri 95% spoľahlivosti ($\alpha=0,05$). Rovnakým spôsobom by sme postupovali aj pri 99 % spoľahlivosti, čiže α by sa rovnala 0,01.



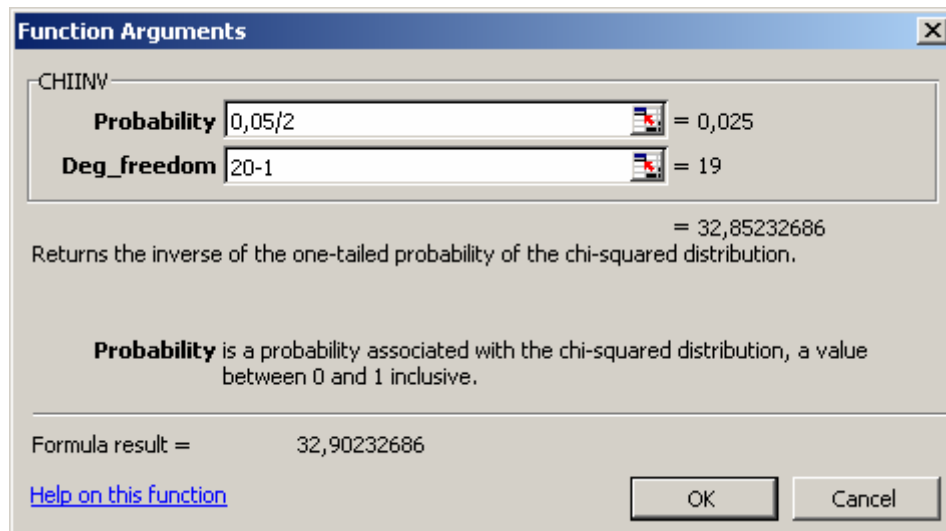
Parametrami funkcie sú hodnoty: $\alpha, \sigma = s_1, n$.

Ďalej je do vzťahu potrebné poznať variabilitu základného súboru. Keďže vychádzame z predpokladu, že základný súbor nepoznáme, odhadneme jeho variabilitu na základe bodového odhadu ($est \sigma = s_1 = 3\,553,69$). Už máme určené všetky hodnoty potrebné do vzťahu, je teda možné dopočítať výslednú hodnotu prípustnej chyby ako aj hranice intervalu spoľahlivosti. Výsledné hodnoty sú uvedené vo výstupe 4.2.. Výsledky budú interpretované súhrnne po dopočítaní ostatných intervalov spoľahlivosti.

Interval spoľahlivosti pre rozptyl a smerodajnú odchýlku

Pri intervale spoľahlivosti pre rozptyl vychádzame z χ^2 rozdelenia. Keďže dané rozdelenie je asymetrické, je potrebné vypočítať dva kvantily pre pravdepodobnosť $\alpha/2$ ako aj $1 - \alpha/2$. Na výpočet použijeme funkciu CHIINV⁴.

⁴ Funkcia CHIINV je prezentovaná len pre pravdepodobnosť $\alpha/2$. Ekvivalentný postup by bol aj v prípade pravdepodobnosti $1 - \alpha/2$



Ako parametre zadávame hodnoty pravdepodobnosti ($\alpha/2$, resp. $1-\alpha/2$) a podobne ako pri Studentovom t-rozdelení druhým parametrom sú stupne voľnosti ($n-1$). Okrem nich je potrebné do vzťahu doplniť ešte hodnotu výberového rozptylu a rozsahu výberového súboru a dopočítat hornú a dolnú hranicu intervalu spoľahlivosti.

Keďže rozptyl nie je možné interpretovať, hranice intervalu spoľahlivosti pre smerodajnú odchýlku vypočítame odmocnením hraníc intervalu spoľahlivosti pre rozptyl.

Interval spoľahlivosti pre podiel

Pri výpočte vychádzame z binomického rozdelenia, ktoré za určitých podmienok je možné aproximovať normálnym rozdelením. Pri praktických výpočtoch je postačujúcou podmienkou

splnenie nasledujúcej nerovnosti: $n > \frac{9}{p(1-p)}$. V takomto prípade vychádzame

z normovaného normálneho rozdelenia a výpočet hodnoty $u_{1-\alpha/2}$ realizujeme pomocou funkcie NORMSINV($1-\alpha/2$, μ , σ).

Pri overovaní normality sme zistili, že nie je splnená potrebná podmienka aproximácie ($20 < 37,5$), preto pri výpočte budeme vychádzať z Fisherovho F rozdelenia. Kvantily Fischerovho rozdelenia určíme pomocou funkcie FINV.

Function Arguments

FINV

Probability 1-0,05/2 = 0,975

Deg_freedom1 2*(8+1) = 18

Deg_freedom2 2*12 = 24

= 0,399568968

Returns the inverse of the F probability distribution: if $p = \text{FDIST}(x, \dots)$, then $\text{FINV}(p, \dots) = x$.

Deg_freedom2 is the denominator degrees of freedom, a number between 1 and 10^{10} , excluding 10^{10} .

Formula result = 0,399568968

[Help on this function](#)

Okrem pravdepodobnosti ($1-\alpha/2$, resp. $\alpha/2$), ďalšími parametrami funkcie sú stupne voľnosti. Stupne voľnosti 1 sa vypočítajú ako $2*(n_1+1)$ a stupne voľnosti 2 ako $2*n_2$. Intervaly spoľahlivosti (p_d, p_h) pre podiel dostaneme potom riešením sústavy dvoch rovníc, ktoré sú uvedené v časti VZORCE.

Vypočítané intervaly spoľahlivosti pre strednú hodnotu, rozptyl, smerodajnú odchýlku a podiel sú uvedené v nasledovnom výstupe.

Výstup 4.2: Výpočet intervalov spoľahlivosti

INTERVALOVÝ ODHAD:

STREDNEJ HODNOTY

1- α	0,95	0,99
α	0,05	0,01
t_α	2,09	2,86
delta	1 663,18	2 273,38
DH	9 881,44	9 271,24
HH	13 207,80	13 818,00

ROZPTYLU

1- α	0,95	0,99
α	0,05	0,01
$\chi^2_{1-\alpha/2}$	8,91	6,84
$\chi^2_{\alpha/2}$	32,85	38,58
DH	7 303 750,79	6 219 055,85
HH	26 940 410,07	35 059 352,55

PODIELU

1- α	0,95	0,99
α	0,05	0,01
aproximácia	37,50	nie je možná

odhad pomocou Fischerovho F rozdelenia

n	20	
n1	8	
n2=N-n1	12	
F 1- $\alpha/2$	0,40	0,29
F $\alpha/2$	2,36	3,12
DH	12*pd/((8+1)*(1-pd))=0,51	12*pd/((8+1)*(1-pd))=0,50
pd	0,23	0,18
HH	12*ph/((8+1)*(1-ph))=2,36	12*ph/((8+1)*(1-ph))=3,12
ph	0,64	0,70

SMERODAJNEJ ODCHÝLKY

1- α	0,95	0,99
α	0,05	0,01
DH	2 702,55	2 493,80
HH	5 190,42	5 921,09

Interpretácia výsledkov:

Na základe intervalu spoľahlivosti pre strednú hodnotu odhadujeme, že priemerné mesačné zárobky pracovníkov sa vo všetkých analyzovaných podnikoch (129) budú pohybovať v intervale od 9 881,44 Sk do 13 207,80 Sk pri 95% pravdepodobnosti (spoľahlivosti), resp. v intervale od 9 271,24 Sk po 13 818 Sk pri 99% pravdepodobnosti. Odhadujeme ďalej, že variabilita v celom súbore (129 podnikov) meraná rozptylom bude v intervale od 7 303 750,79 Sk² do 26 940 410,07 Sk² pri 95% spoľahlivosti, resp. od 6 219 055,85 Sk² do 35 059 352,55 Sk² pri 99% spoľahlivosti. Odhadujeme, že kolísanie priemerných mesačných zárobkov pracovníkov v celom súbore merané smerodajnou odchýlkou sa bude pohybovať v intervale od 2 702,55 Sk do 5 190,42 Sk (95% pravdepodobnosť), resp. od 2 493,80 Sk do 5 921,09 Sk (99% pravdepodobnosť). Vychádzajúc z výberového podielu odhadujeme, že podiel družstiev v celom súbore sa bude nachádzať v intervale od 23% do 64% pri spoľahlivosti 95%, resp. v intervale od 18% do 70% pri spoľahlivosti 99%.

Hoci je pri intervalových odhadoch zaručená vyššia spoľahlivosť, ich nevýhodou je to, že poznáme len interval, v rámci ktorého sa bude odhadovaný parameter pohybovať, t.j. nepoznáme presne ktorá hodnota v rámci neho to bude. Dokonca pri zvyšovaní spoľahlivosti (z 95% na 99%), ako je možné vidieť aj z výpočtov, sa intervaly spoľahlivosti rozširujú, čo je pochopiteľné, pretože plocha pod krivkou predstavuje už 99%, t.j. dolná hranica sa zníži a horná hranica zvýši. Na druhej strane sa však pri zvyšovaní spoľahlivosti znižuje presnosť odhadov. Keďže intervalový odhad je závislý nielen od zvolenej spoľahlivosti, ale aj od ďalších hodnôt (rozsahu výberového súboru, variability základného súboru), je vhodné zvyšovať presnosť intervalového odhadu práve cez tieto hodnoty.