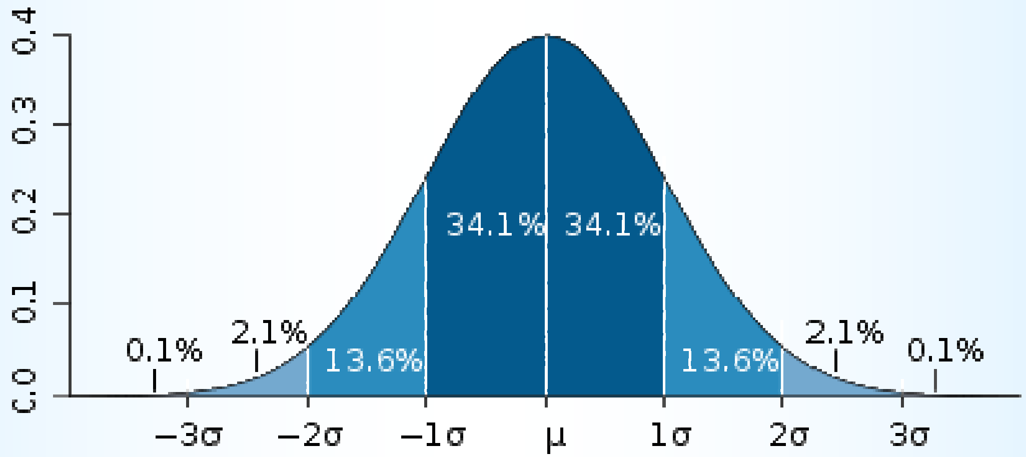
VII. Centrální limitní teorém a Bodové odhady

Centrální limitní teorém 1

Normální rozdělení



je výjimečné mezi ostatními rozděleními spojitých NV také tím, že je rozdělením "limitním", k němuž se jiná rozdělení „blíží“ s rostoucím rozsahem výběru

Centrální limitní teorém 2

Normálním rozdělením, jako rozdělením limitním, se zabývá:

centrální limitní teorém

- má různé formy v závislosti na podmínkách, jejichž splnění se požaduje. Nejdůležitější formou je

věta Lindbergova – Lévyho - má dvě varianty: pro **průměr** a pro **úhrn** (součet).

Centrální limitní teorém 3

Lindberg - Lévyho věta pro průměr:

Průměr výsledků z dostatečně velkého počtu ($n > 30$) realizací stejné náhodné veličiny s libovolným rozdělením pravděpodobnosti (střed. hod. μ, rozptyl σ^2) má „přibližně“ normální rozdělení pr-sti:

$$N(\mu, \frac{\sigma^2}{n})$$

Příklad 4: Platby faktur

Z populačního souboru 1250 faktur (od 100 do 10 tis. Kč) byl vybrán vzorek o 50 fakturách. Všech možných vzorků je

$$\binom{1250}{50} = 2 \cdot 10^{91} = 2000\ldots0 \text{ (91 nul!!!)}$$

$$\boxed{\mu = 5097 \quad \sigma = 412{,}5}$$

Parametry populačního souboru - neznáme

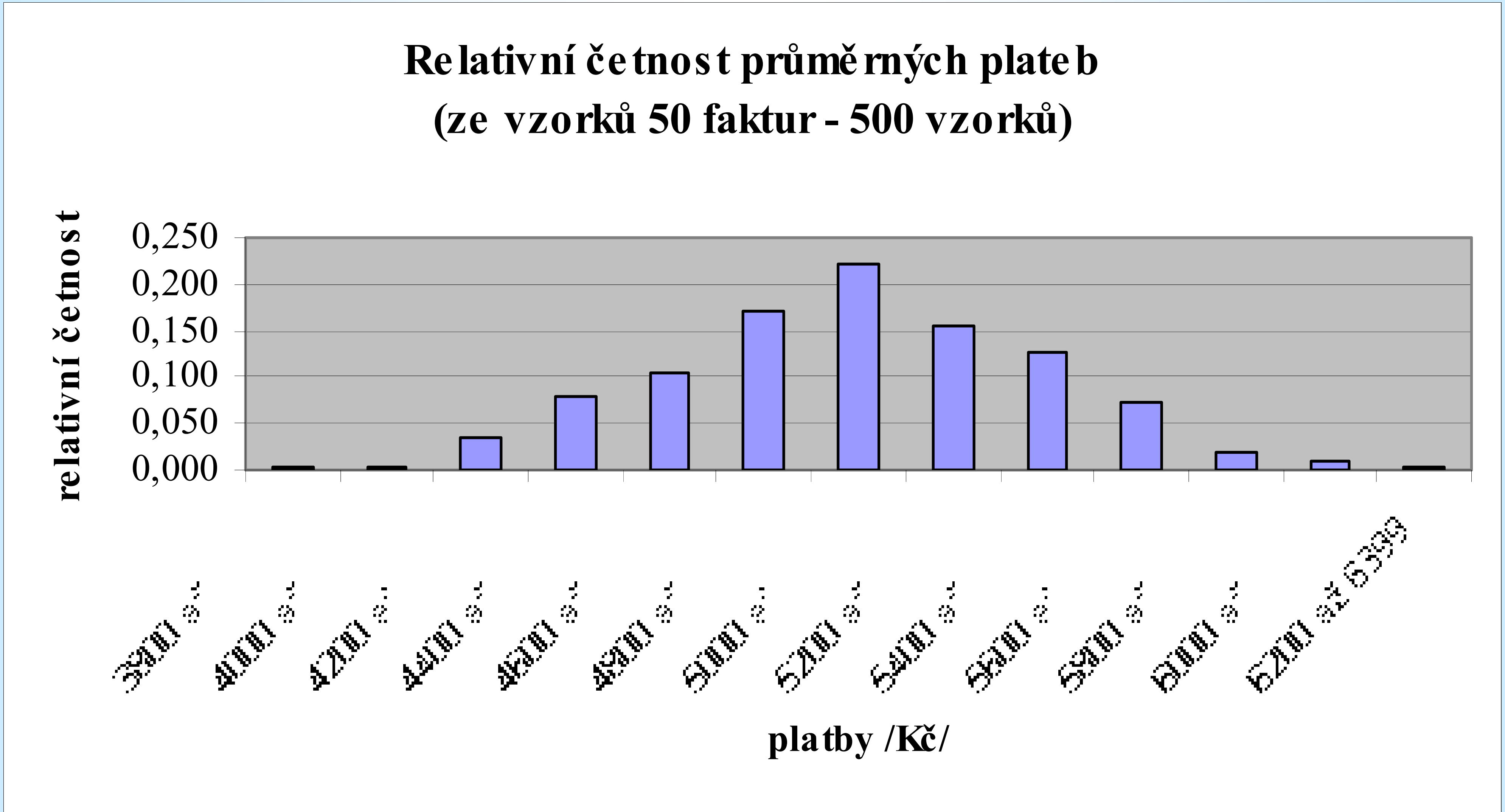
Průměrné platby vzorků se pohybují mezi 3800 až 6400

Histogram četnosti

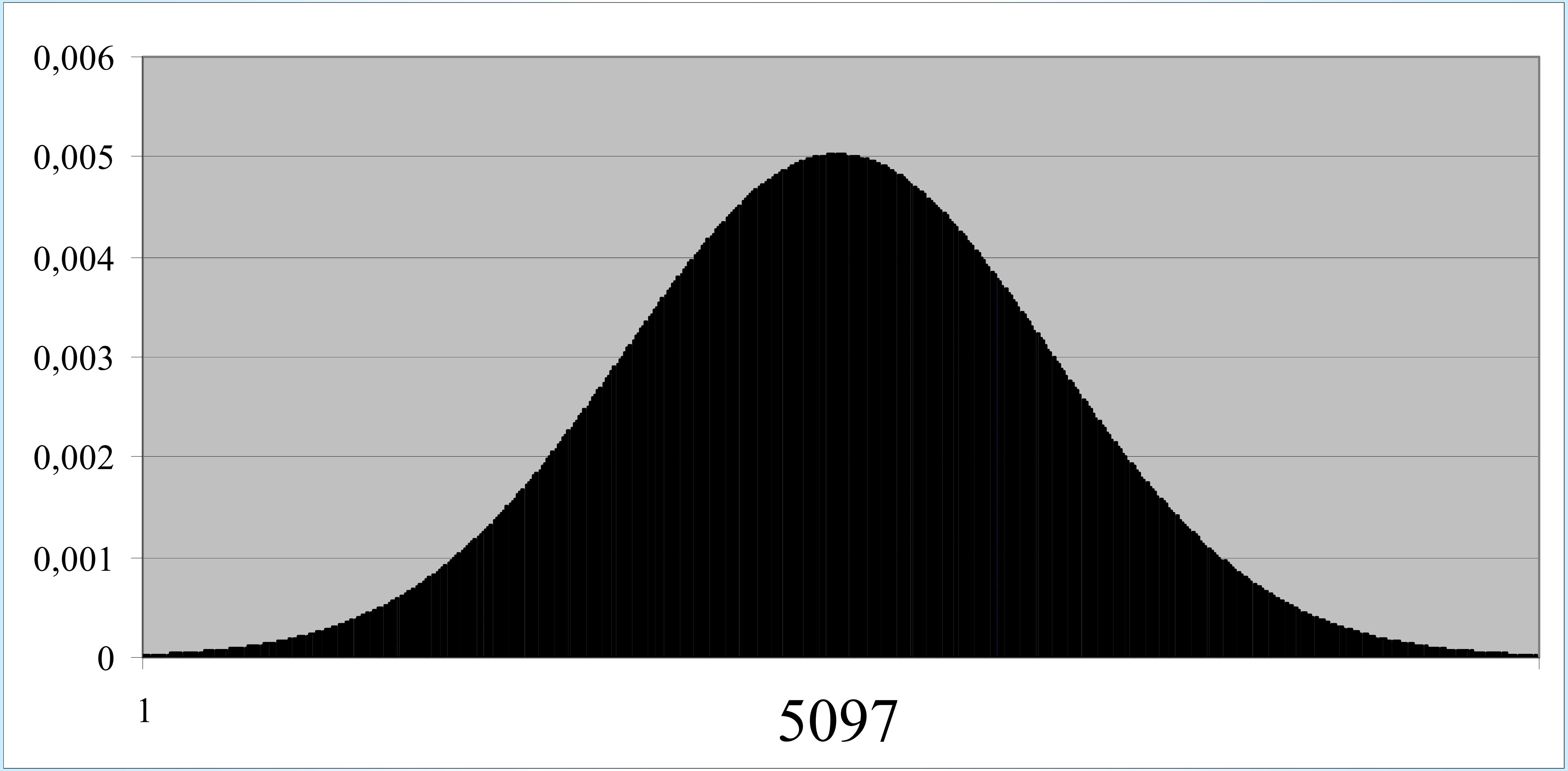
pro 500 vzorků po 50 fakturách

Interval platby	četnost		relativní četnost
3800 až 3999	1	1/500 =	0,002
4000 až 4199	2	2/500 =	0,004
4200 až 4399	17	17/500 =	0,034
4400 až 4599	39	39/500 =	0,078
4600 až 4799	52	52/500 =	0,104
4800 až 4999	85	85/500 =	0,170
5000 až 5199	110	110/500 =	0,220
5200 až 5399	77	77/500 =	0,154
5400 až 5599	64	64/500 =	0,128
5600 až 5799	37	37/500 =	0,074
5800 až 5999	10	10/500 =	0,020
6000 až 6199	4	4/500 =	0,008
6200 až 6399	2	2/500 =	0,004
Celkem	500		1,000

Histogram četnosti průměrných plateb z 50 faktur



Histogram četnosti průměrných plateb 50 faktur - 5000 vzorků



Výběrové šetření, výběrový plán

- V praktických úlohách bývají populační soubory rozsáhlé nebo nejednoznačně vymezené!
- ***Výběrové šetření*** *(VŠ)* - proces získání výběrového vzorku
- VŠ slouží k tomu, abychom získali vzorek se strukturou co možná nejvíce podobnou populačnímu souboru
- ***Výběrový plán*** - způsob nebo metoda „generování“ prvků výběru

Hlavními typy výběrových plánů jsou:

- ***Anketa*** (charakteristika: ochota odpovídat)
- ***Záměrný výběr*** (kvótní výběr, systematický výběr)
- ***Náhodný výběr*** - nejdůležitější typ (všechny prvky mají stejnou pr-st dostat se do výběru)

Výběrové šetření - příklad

- ***Populační soubor***: zákazníci supermarketu TESCO v Karviné v roce 2013
- ***Sledovaný statistický znak:*** hodnota nakoupeného zboží (Kč)
- ***Výběrový plán*** - způsob nebo metoda „generování" prvků ***výběrového souboru***, velikost vzorku?

Hlavními typy výběrových plánů jsou:

- ***Anketa*** (dotazník vyplní zákazník u východu – podle ochoty vyplnit)
- ***Záměrný výběr*** (kvótní výběr: 65% ženy, 35% muži)
- ***Náhodný výběr*** - nejdůležitější typ (všechny prvky mají stejnou šanci dostat se do výběru – např. podle tabulky náhodných čísel)

Náhodný výběr (NV)

- *NV* - soubor vzájemně nezávislých realizací náhodné veličiny
- Každá z jednotek populačního souboru má stejnou šanci (pravděpodobnost) dostat se do výběru
- ***Prostý** NV* - jednotky do vzorku jsou vybírány přímo v jediném kroku, např. losováním
- ***Vícestupňový** NV* (výběr v několika krocích)
- Prostý náhodný výběr se realizuje **2 způsoby**:

  s vracením a bez vracení
- - ***výběr bez vracení*** (bez opakování) – vybraná jednotka se do populačního souboru již nevrací, může být tedy vybrána nejvýše jedenkrát
- - ***výběr s vracením*** (opakováním) - vybraná jednotka se do populačního souboru vrací, může být tedy vybrána vícekrát.
- V praxi se častěji používá **výběr bez vracení**

Princip bodového a intervalového odhadu

Problém: zjistit (odhadnout) hodnoty charakteristik **populačního** souboru, např. μ nebo σ^2, pomocí charakteristik **výběrového** souboru, např. $\bar{x}$ a s^2

Příklad: Máme odhadnout velikost průměrné hodnoty nákupů μ u zákazníků supermarketu TESCO Ka v roce 2013 s pravděpodobností **95%** Byl vybrán náhodný vzorek **64** zákazníků. Z jejich nákupů jsme zjistili průměr $\bar{x}$ = **450** Kč a směrodatnou odchylku s = **128** Kč.

Řešení:

Průměrnou velkost nákupů populačního souboru (všech) zákazníků TESCO odhadneme jako interval [L,P]: L je levý krajní bod, P je pravý krajní bod. Střed intervalu stanovíme jako $\overline{x}$. Hledaný interval:

95% kvantil $N(0,1)$

$$L = \overline{x} - \Delta, \quad P = \overline{x} + \Delta$$

Přitom

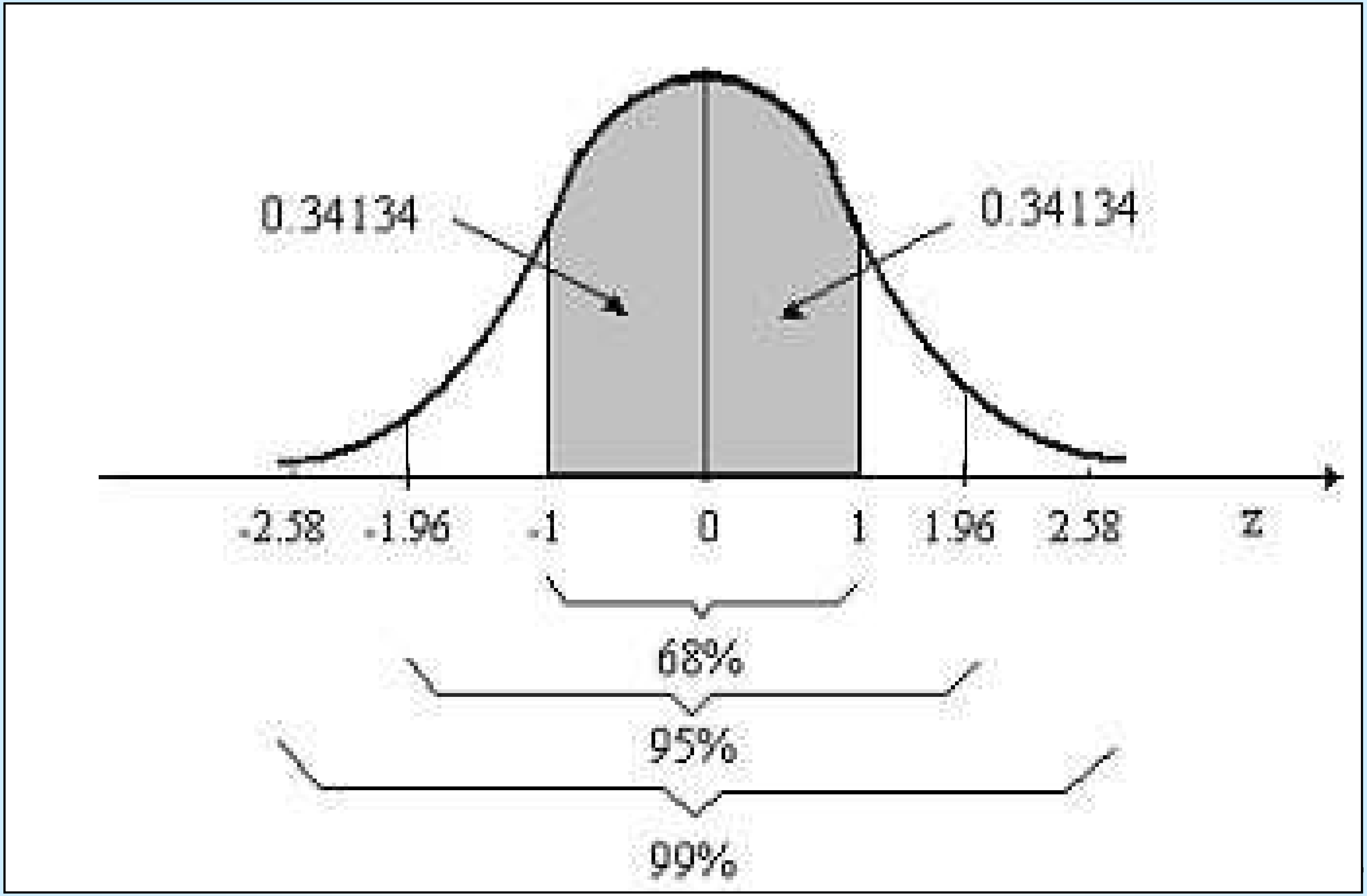
$$\Delta = 1{,}96 \frac{s}{\sqrt{n}}$$

Kde $n = 64$, $s = 128$, tj. $\Delta = 1{,}96 \dfrac{128}{\sqrt{64}} \approx 31$

Hledaný interval: [L,P] = [450-31,450+31] = [419,481], tedy:

Neznámý populační průměr μ leží v **intervalu spolehlivosti** [419,481] s pravděpodobností 95 procent (téměř jistota)

Významné hodnoty normovaného normálního rozdělení $N(0,1)$



Bodové odhady parametrů μ, σ^2 a p

Dvojí význam **bodového odhadu** (BO):

- jako ***statistika*** (funkce náhodné veličiny – předpis k výpočtu čísla, tj. realizace náhodné veličiny)
- jako konkrétní ***číslo***, které se tím předpisem vypočte

Příklad. Mějme náhodnou veličinu $X = X_j$ s neznámým parametrem μ střední hodnoty,

$x_1, x_2, \ldots, x_n$ jsou konkrétní hodnoty - realizace náhodné veličiny,

tj. náhodný výběr z X .

Výběrový průměr: $\overline{X} = \dfrac{X_1 + \ldots + X_n}{n}$ - statistika = BO

Realizací statistiky výběrový průměr je číslo: $\overline{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

což je bodový odhad parametru μ

3 kritéria „dobrého“ bodového odhadu

- ***Nestrannost*** -střední hodnota zvolené statistiky (BO) = správné hodnotě odhadovaného parametru. Statistika splňující uvedený požadavek je *nestranný (nezkreslený, nevychýlený) odhad parametru*
- ***Konzistence*** -s rostoucím rozsahem výběru *n* roste pravděpodobnost, že zvolená statistika poskytne hodnotu bližší skutečné hodnotě parametru. Takovou statistiku (BO) nazýváme *konzistentním odhadem parametru*

\- Aby byl BO konzistentní, stačí, aby byl **nestranný** a navíc byla splněna podmínka, že **limita rozptylu BO (statistiky) je rovna nule** při rozsahu výběru zvětšujícím se nade všechny meze

- ***Vydatnost*** *(eficience)* -statistika s nejmenším rozptylem se nazývá *vydatný* (nebo také *eficientní*) *odhad parametru*

Bodový odhad μ a σ^2

- Statistika $\overline{X}$ (aritmetický průměr) je BO μ (střední hodnoty), protože platí

$$E(\overline{X}) = \mu \qquad Var(\overline{X}) = \frac{\sigma^2}{n}$$

jde o nestranný a konzistentní BO parametru μ

Jiné *možné BO* parametru μ : $\dfrac{\max x_i - \min x_i}{2}$, $Med\,(X)$???

- Statistika $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$!!!

je nestranný a konzistentní BO parametru σ^2

Příklad:

- Populační soubor (výsledky 5 testů):
  33, 12, 24, 40, 37

Populace:					průměr	rozptyl
33	12	24	40	37	29,2	128,70
Vzorky (po 3):					průměr	rozptyl
33	12	24			23,00	111,00
33	12	40			28,33	212,33
33	12	37			27,33	180,33
12	24	40			25,33	197,33
12	24	37			24,33	156,33
24	40	37			33,67	72,33
33	24	40			32,33	64,33
33	24	37			31,33	44,33
33	40	37			36,67	12,33
12	40	37			29,67	236,33
				Průměr:	29,20	128,70