Induktivní statistika

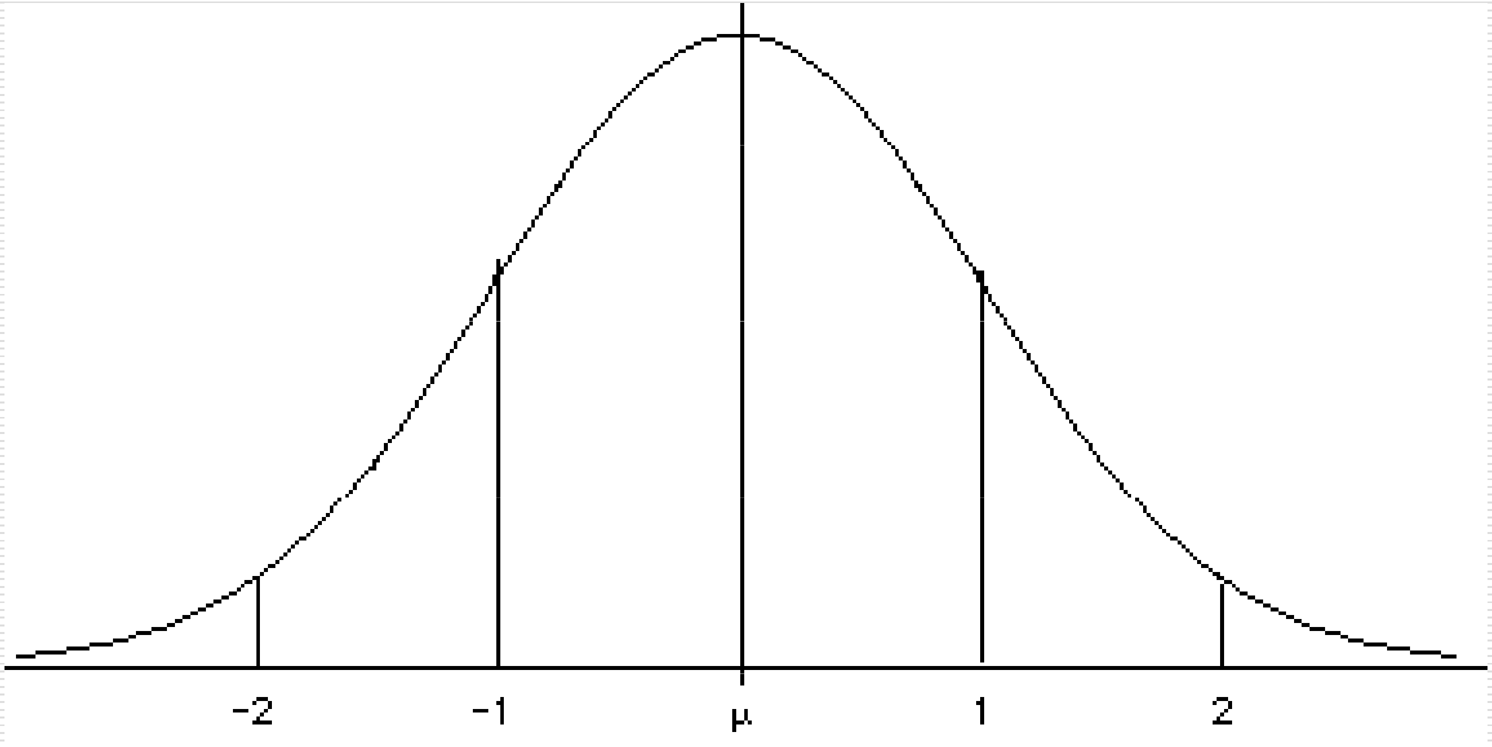
- normální rozdělení
- rozdělení výběrových průměrů

Normální rozdělení

- normální rozdělení je symetrické, unimodální, zvonovitého tvaru
- označuje se i jako Gaussova křivka

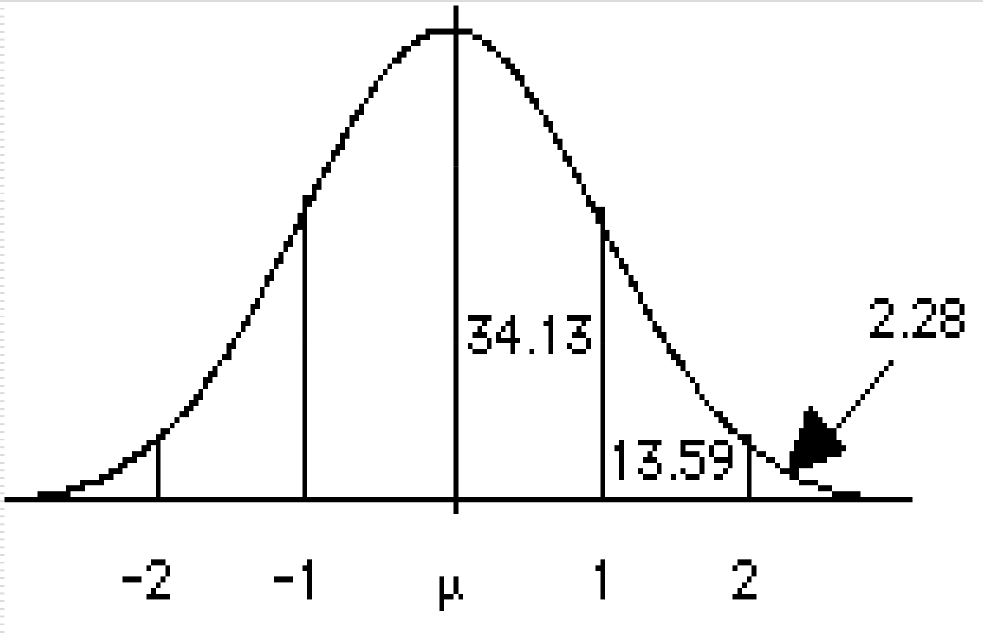
$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2}$$

Normální rozdělení



Normální rozdělení

- 34.13% skórů spadá mezi průměr a 1 směr. odchylku
- 13.59% hodnot spadá mezi 1. a 2. směr. odchylku
- 2.28% hodnot spadá nad 2. směr. odchylku

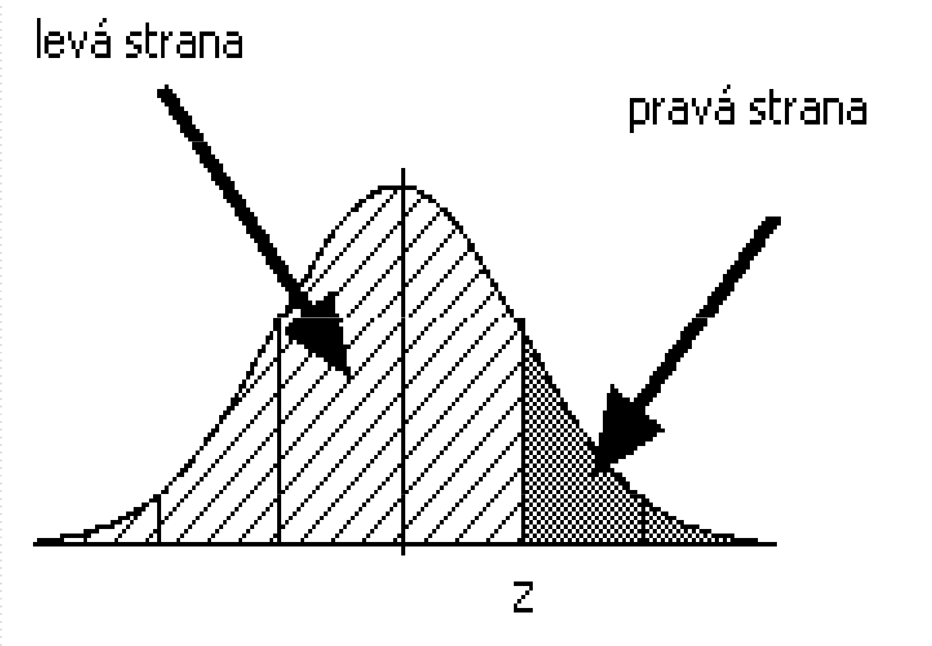


Normální rozdělení

- tabulka normálního rozdělení
(z rozdělení)
- důležitý nástroj, obvykle jako apendix
v učebnicích statistiky
(spolu s dalšími tabulkami)
- umožňuje zjistit hustotu oblasti pod
křivkou (tj. pravděpodobnost)
pro jednotlivé z-skóry

Normální rozdělení

z	levá strana	pravá strana
0.00	0.5000	0.5000
0.01	0.5040	0.4960
...	...	...
0.30	0.6179	0.3821
0.31	0.6217	0.3783
...	...	...
1.00	0.8413	0.1587
...	...	...



Normální rozdělení - příklady

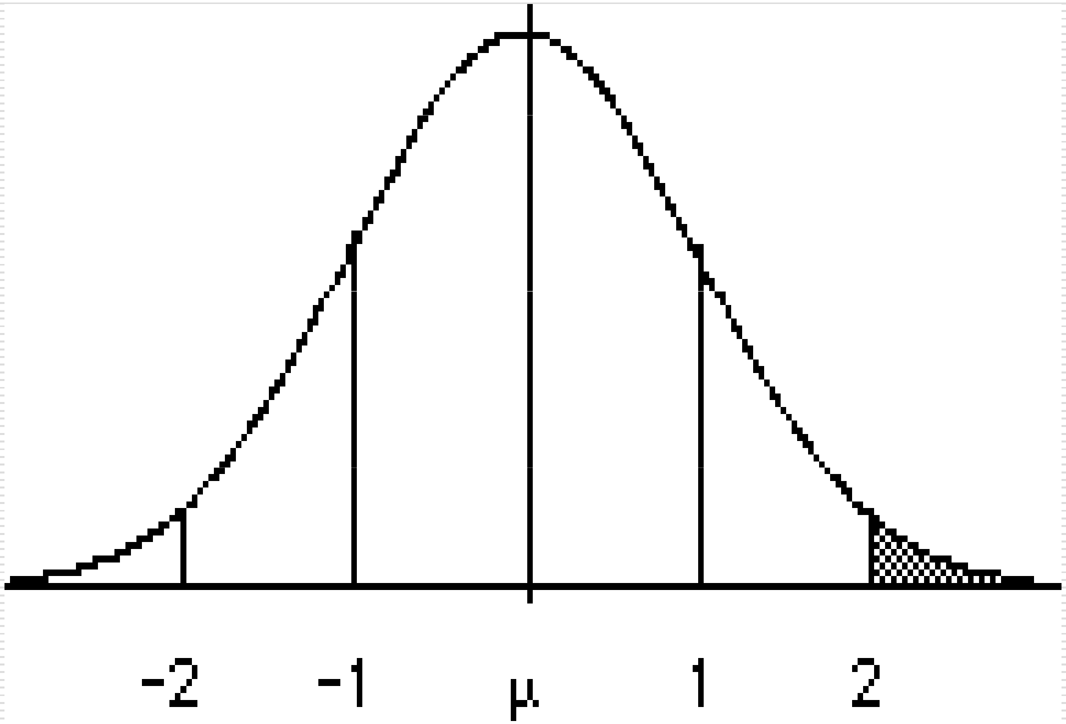
- postup při zjišťování pravděpodobnosti z tabulky:
  - načrtnout si normální rozdělení, s hodnotou průměru a směr. odch.
  - zakreslit hledanou hodnotu (v přibližné vzdálenosti od průměru), vystínovat hledanou oblast
  - převést hodnotu X na z-skór
  - najít v tabulce pravděpodobnost

Normální rozdělení - příklady

- Jaká je pravděpodobnost, že náhodně vybraná osoba z populace bude mít IQ 130 nebo vyšší? (μ = 100, σ =15)

Normální rozdělení - příklady

- z = 2



Normální rozdělení - příklady

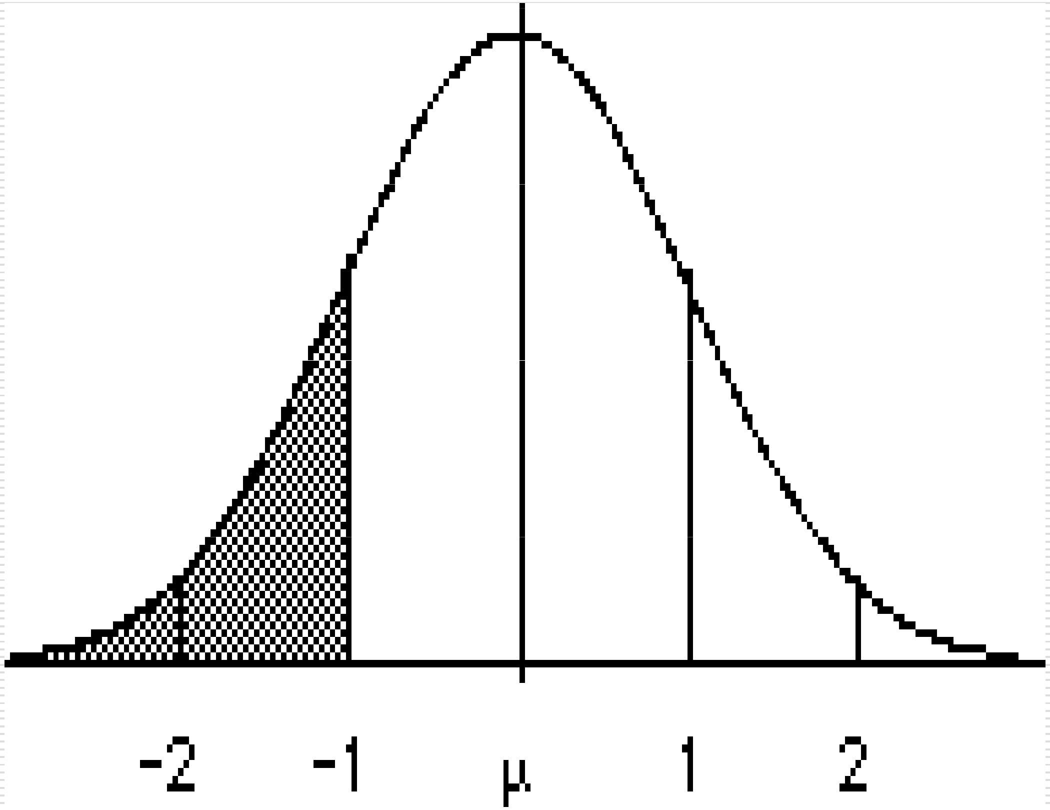
- Jaká je pravděpodobnost, že náhodně vybraná osoba z populace bude mít IQ 130 nebo vyšší?
- $z = 2$
- $p = 1 - (0.50 + 0.4772) = 0.0228$
  tj. **2,3%**

Normální rozdělení - příklady

- Jaká je pravděpodobnost, že náhodně vybraná osoba z populace bude mít IQ 85 nebo nižší?

Normální rozdělení - příklady

- z = -1



Normální rozdělení - příklady

- Jaká je pravděpodobnost, že náhodně vybraná osoba z populace bude mít IQ 85 nebo nižší?
- $z = -1$
- $p = 1 - (0.50+0.3413) = 0.1587$ tj. **15,9%**

Normální rozdělení - příklady

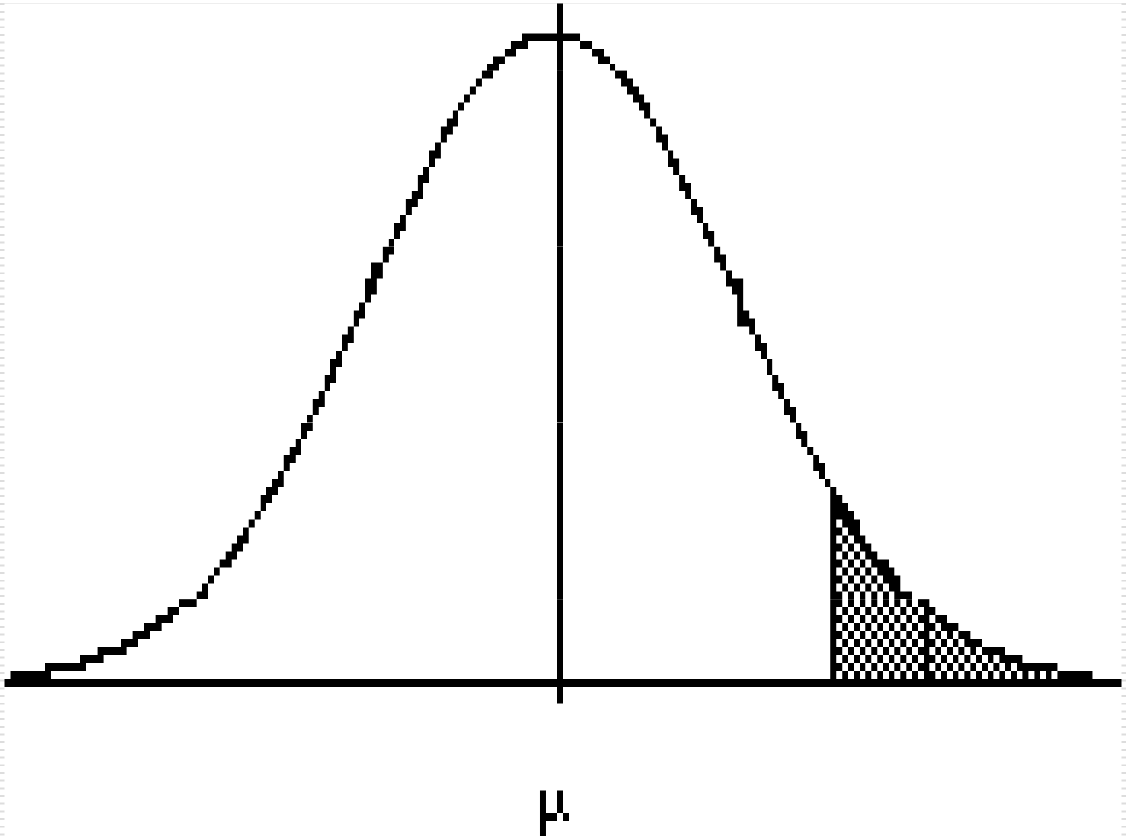
- postup při zjišťování z-skóru z tabulky:
  - načrtnout si normální rozdělení
  - vystínovat oblast odpovídající zadané pravděpodobnosti
  - v tabulce vyhledat příslušný z-skór
  - vypočítat z něj hrubý skór

Normální rozdělení - příklady

- Jakou minimální hodnotu IQ musí člověk mít, aby patřil mezi 5% osob s nejvyššími hodnotami IQ?

Normální rozdělení - příklady

- p = 0.05



Normální rozdělení - příklady

- Jakou minimální hodnotu IQ musí člověk mít, aby patřil mezi 5% osob s nejvyššími hodnotami IQ?
- p = 0.05
- z tabulky (hledáme hodnotu nejbližší 0.50-0.05, tj. 0.45): **z = 1.65**
- X = (1.65)x(15) + 100 = **124.75**

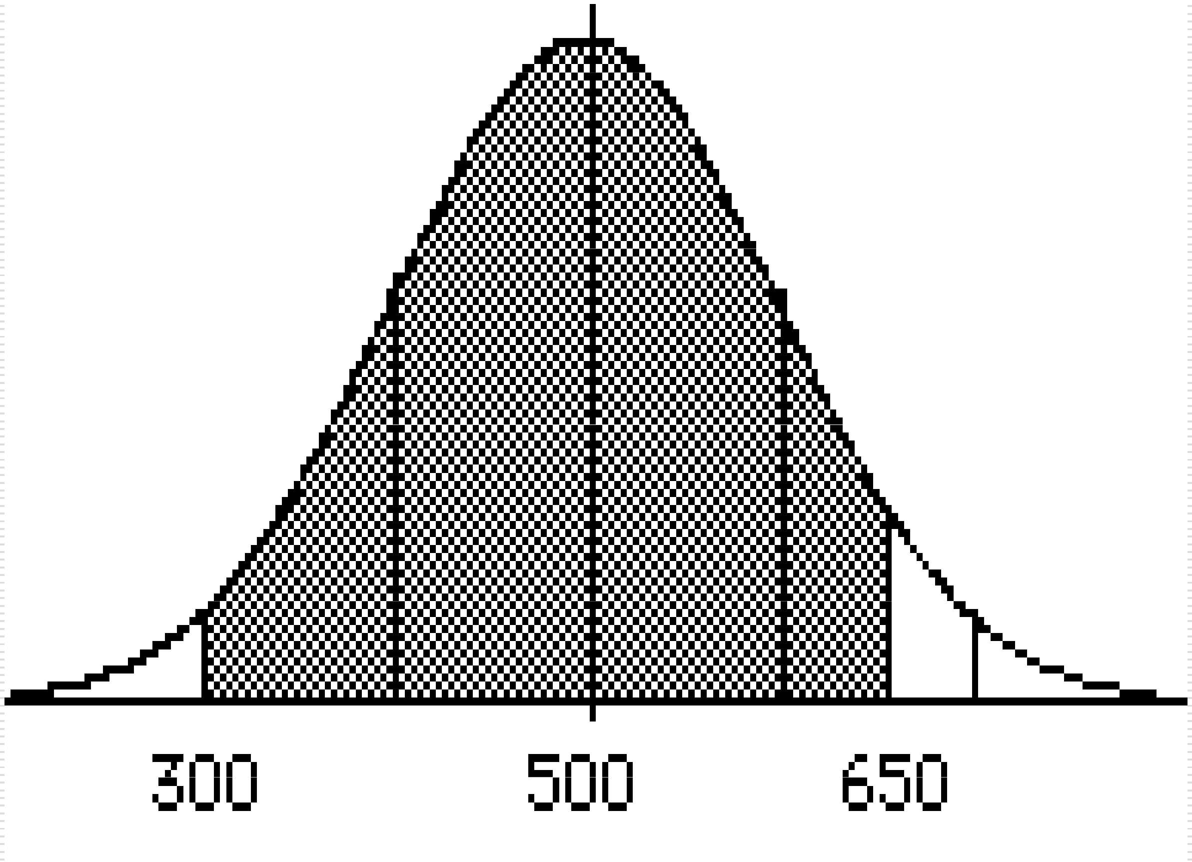
Normální rozdělení - příklady

- někdy chceme zjistit pravděpodobnost, že skór bude spadat do určitého intervalu
- postup:
  - načtrtnout graf a vystínovat zadanou oblast
  - oba (ohraničující) skóry převést na z-skóry
  - vyhledat pravděpodobnosti < nebo > skóru
  - sečíst či odečíst pravděpodobnosti

Normální rozdělení - příklady

- Jaká je pravděpodobnost, že náhodně vybraný student bude v testu z psychologie skórovat mezi 300 a 650 body? (μ = 500, σ =100)

Normální rozdělení - příklady



Normální rozdělení - příklady

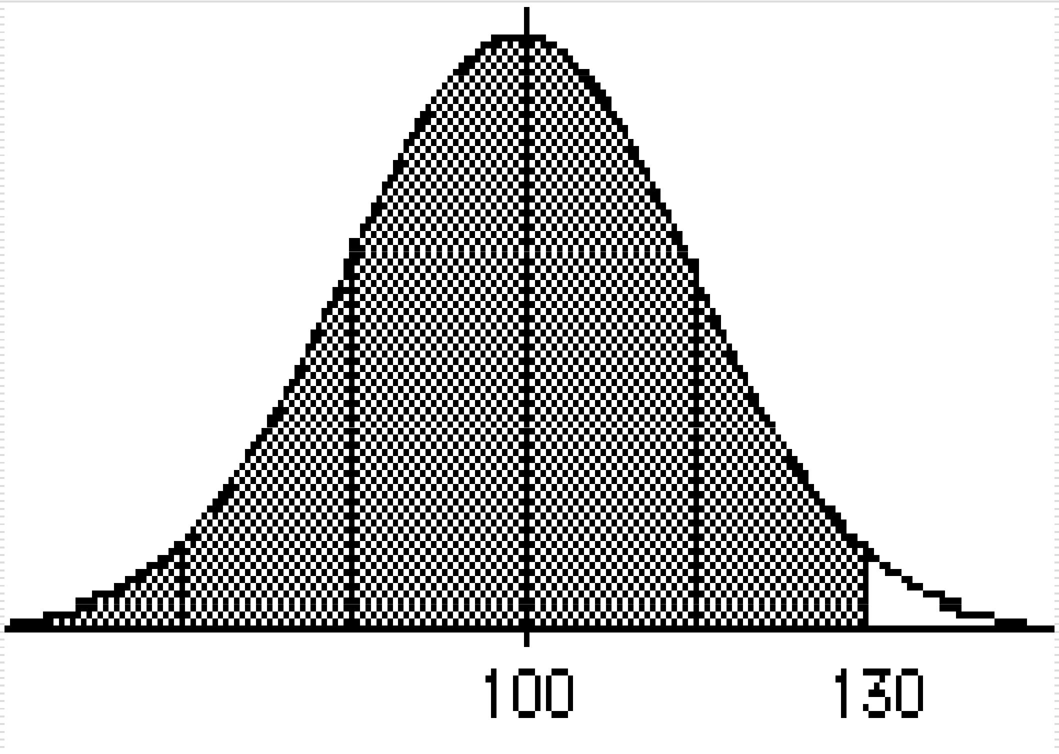
- Jaká je pravděpodobnost, že náhodně vybraný student bude v testu z psychologie skórovat mezi 300 a 650 body? ($\mu = 500$, $\sigma = 100$)
- $p(300 < x < 650) = 0.4772 + 0.4332 = \mathbf{0.9104}$

Normální rozdělení - příklady

- příklad: kolik procent osob má nižší hodnoty IQ než člověk s IQ 130?

Normální rozdělení - příklady

- $z = 2$



Normální rozdělení - příklady

- Kolik procent osob má nižší hodnoty IQ než člověk s IQ 130?
- z tabulky: pro z = 2

p = 0.50 + 0.4772 = 0.9772

97.72% osob má nižší skór

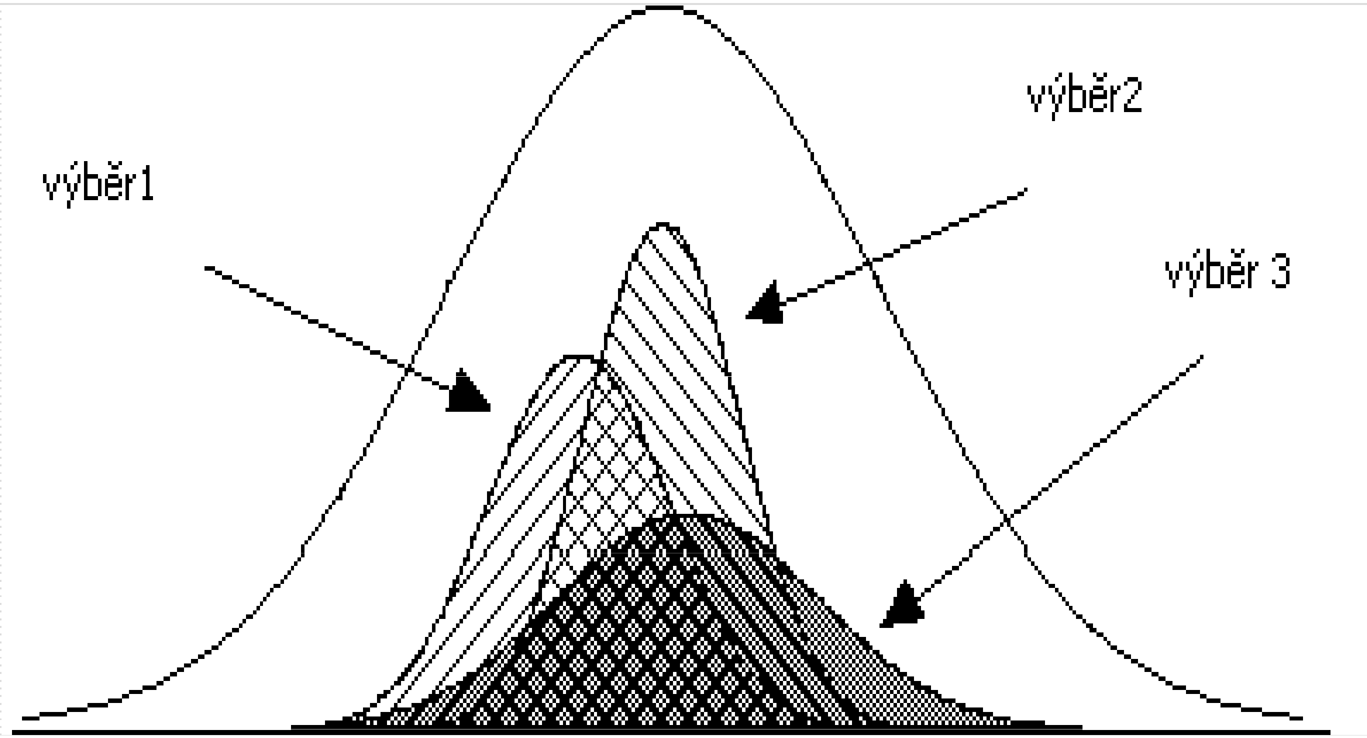
Rozdělení výběrových průměrů

- cílem induktivní statistiky je odhadnout parametry populace z charakteristik vzorku (výběrového souboru)
- např. odhadem průměru populace bude průměr vzorku
- odhad je vždy zatížen určitou **výběrovou chybou**

Rozdělení výběrových průměrů

- předpokládejme, že z jedné populace vybereme 3 různé vzorky
- budou se nejspíš navzájem lišit ve tvaru rozdělení hodnot, průměru i variabilitě
- jak se rozhodneme, který z nich zvolit pro odhad průměru populace ??

Rozdělení výběrových průměrů



Rozdělení výběrových průměrů

- pokud bychom spočítali průměry ze všech možných výběrů o určité velikosti *n*, budou tvořit tzv. **rozdělení výběrových průměrů** (sampling distribution)

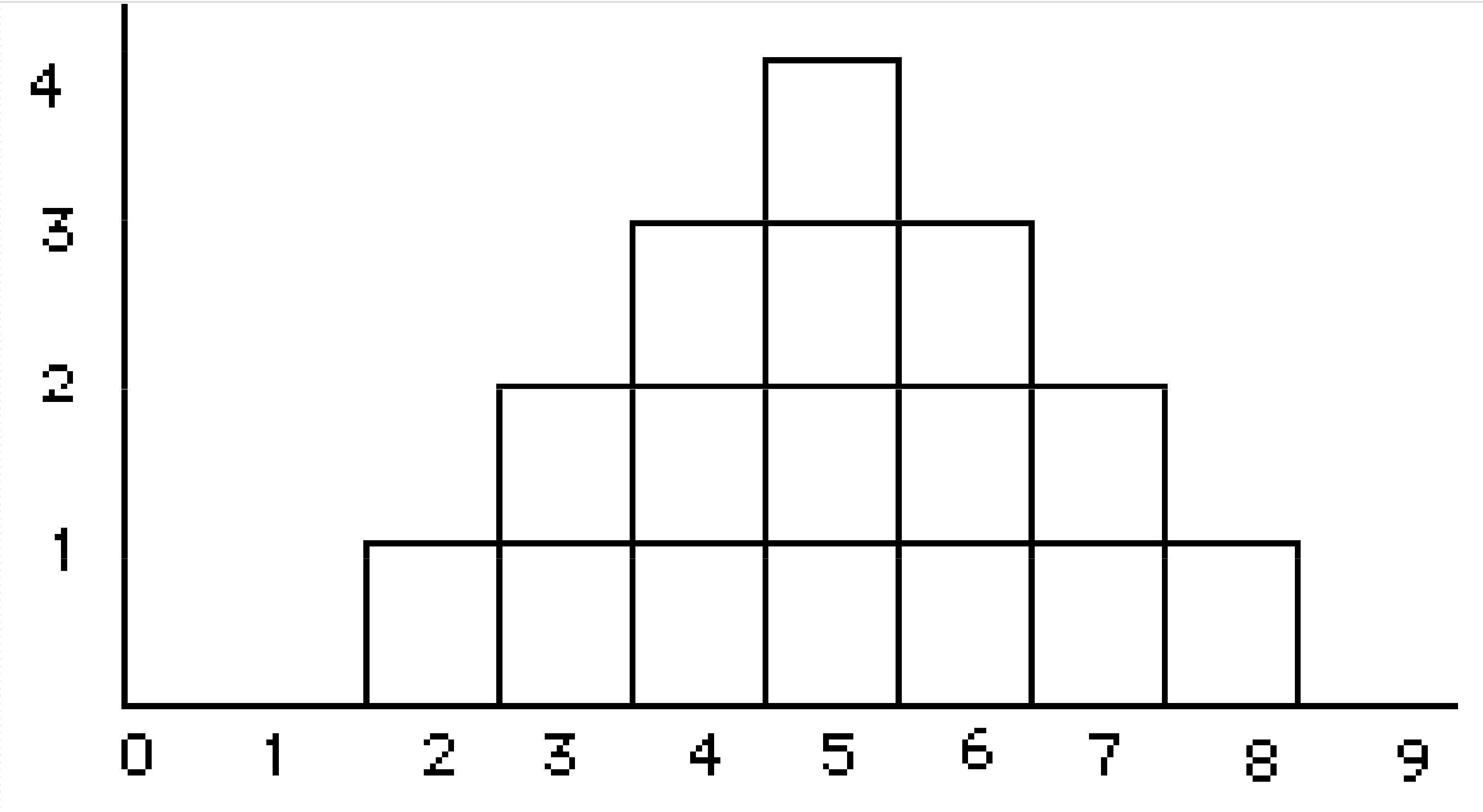
Rozdělení výběrových průměrů

- **příklad**: populace hodnot 2, 4, 6, 8
- průměr $\mu = 5$
- předpokládejme, že průměr neznáme a pokoušíme se ho odhadnout ze vzorku n=2
- v tabulce jsou uvedeny všechny možné výběrové soubory

Rozdělení výběrových průměrů

výběr	první skór	druhý skór	průměr vzorku
1	2	2	2
2	2	4	3
3	2	6	4
4	2	8	5
5	4	2	3
6	4	4	4
7	4	6	5
8	4	8	6
9	6	2	4
10	6	4	5
11	6	6	6
12	6	8	7
13	8	2	5
14	8	4	6
15	8	6	7
16	8	8	8

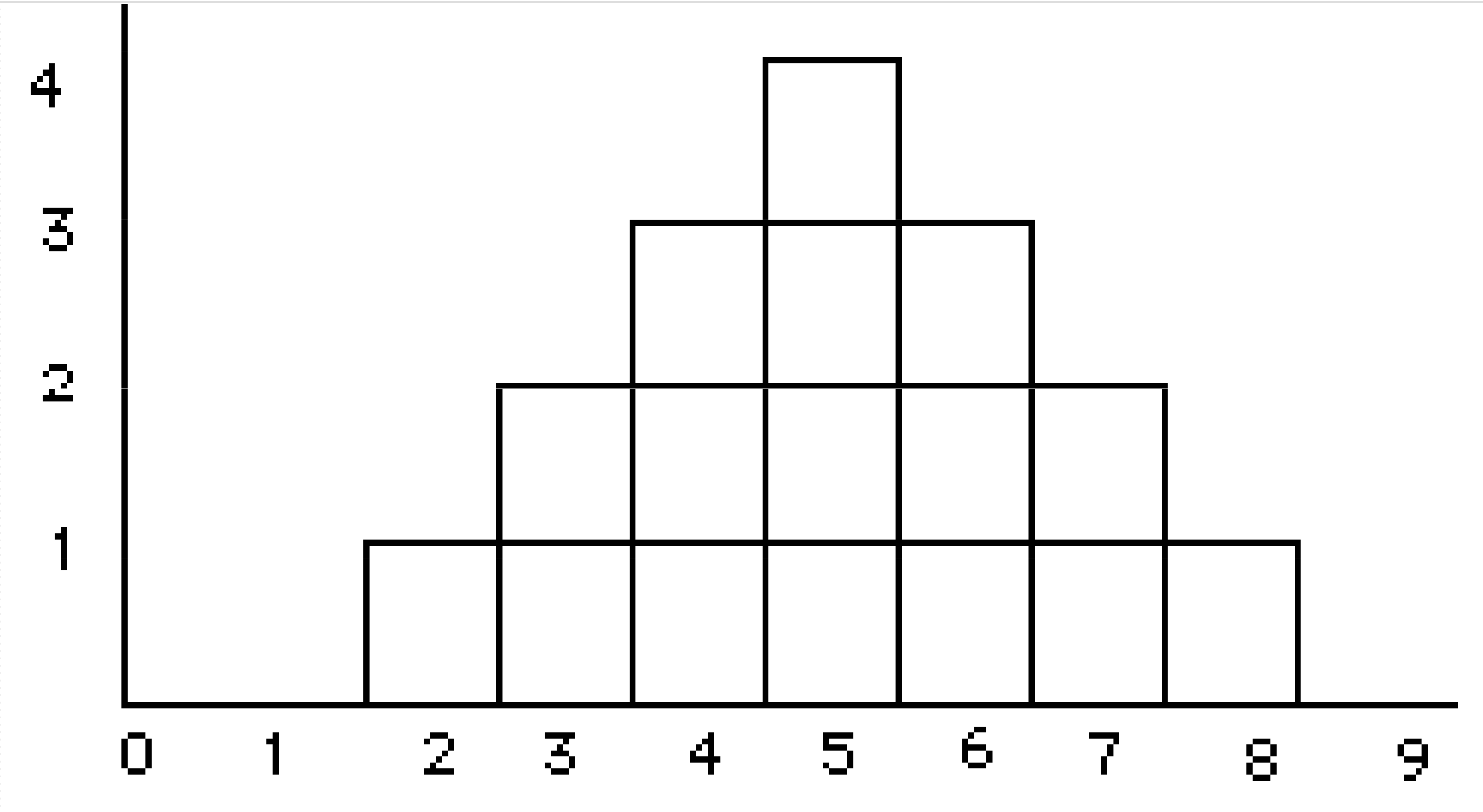
Rozdělení výběrových průměrů



Rozdělení výběrových průměrů

- jaká je pravděpodobnost, že z této populace vybereme vzorek s průměrem vyšším než 7?

Rozdělení výběrových průměrů



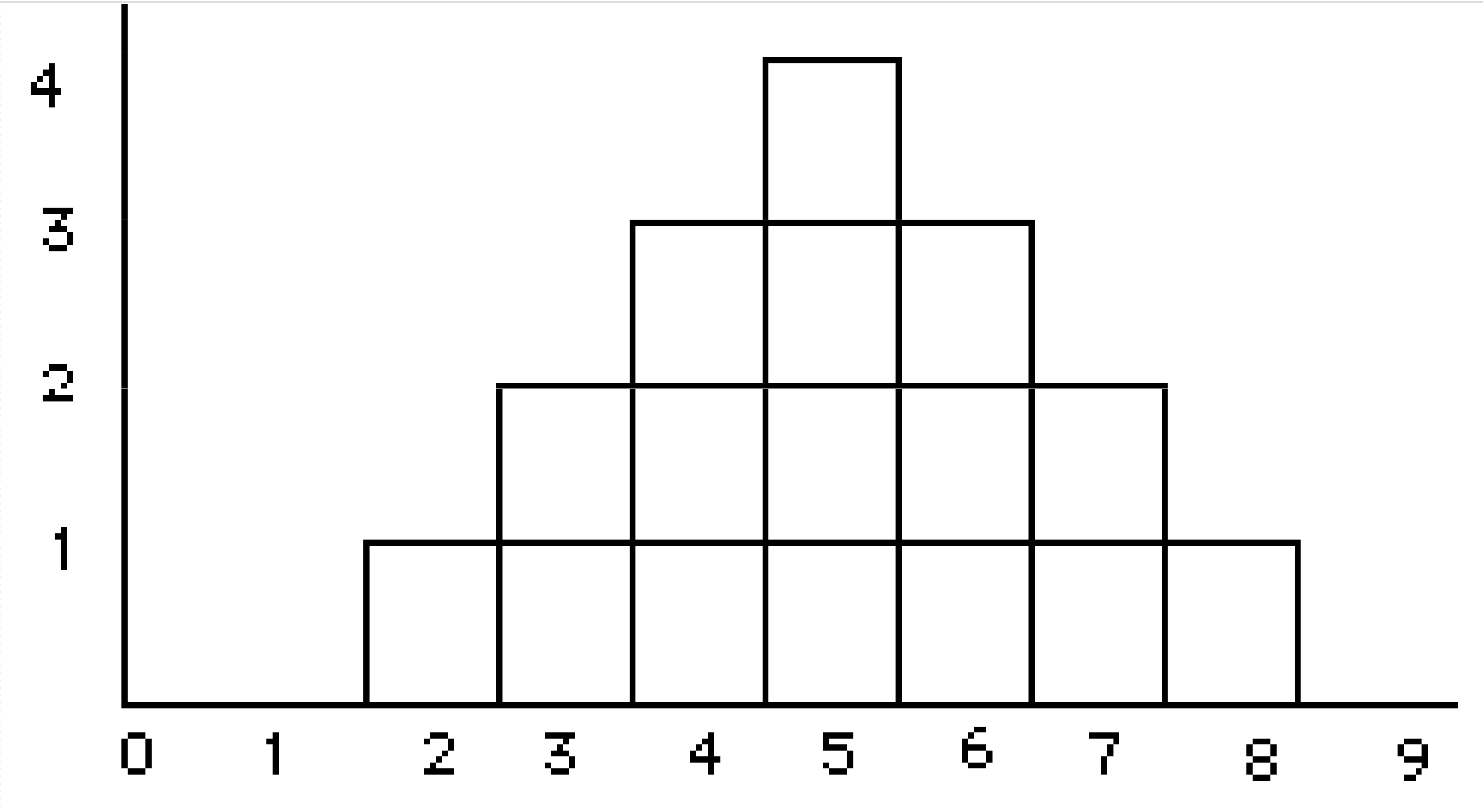
Rozdělení výběrových průměrů

- jaká je pravděpodobnost, že z této populace vybereme vzorek s průměrem vyšším než 7?
- v rozdělení výběrových průměrů je takový vzorek jen 1 ze 16 – tj. pravděpodobnost takového vzorku je 1/16 = 0.0625, tj. 6%

Rozdělení výběrových průměrů

- jaká je pravděpodobnost, že náhodně vybraný vzorek 2 čísel z této populace bude mít průměr roven průměru populace, tj. 5?

Rozdělení výběrových průměrů



Rozdělení výběrových průměrů

- jaká je pravděpodobnost, že náhodně vybraný vzorek 2 čísel z této populace bude mít průměr roven průměru populace, tj. 5?
- tato pravděpodobnost je 4/16, tj. 25%

Rozdělení výběrových průměrů

- většina populací i vzorků je mnohem větší
- ale existují určité základní vlastnosti rozdělení výběrových průměrů (RVP)
- **tvar** – RVP se při dostatečně velkém vzorku (>30 případů) blíží **normálnímu rozdělení**

Rozdělení výběrových průměrů

- **průměr** tohoto rozdělení (=průměr průměrů všech teoretických výběrů) je roven **průměru populace**
- označuje se také jako očekávaná hodnota průměru vzorku

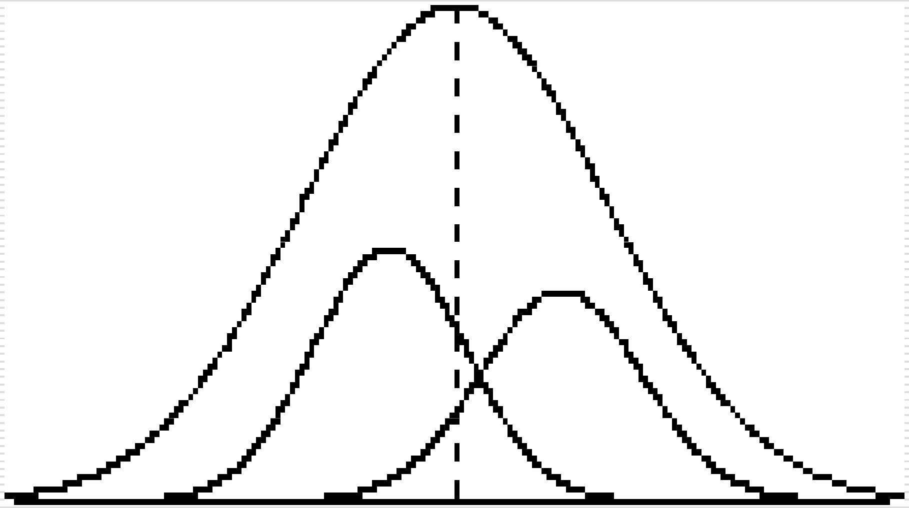
Rozdělení výběrových průměrů

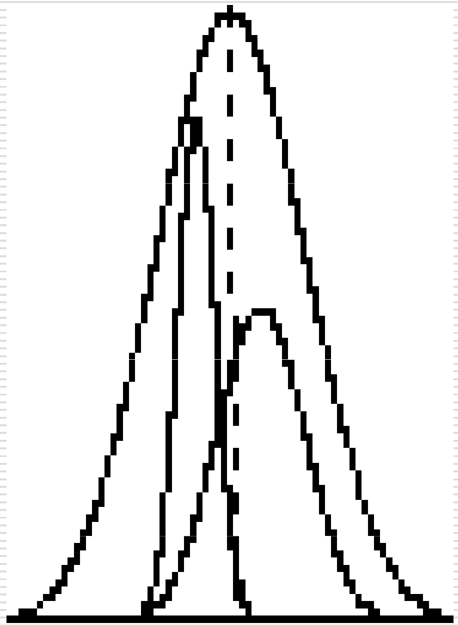
- **variabilita** – směrodatná odchylka RVP se označuje jako výběrová nebo standardní/směrodatná chyba průměru (standard error)
- jde o **směrodatnou odchylku výběrových průměrů** <u>od průměru populace</u>
- ukazuje, jak spolehlivý je odhad populačního průměru z průměru vzorku – tj. jak velkou chybou je odhad zatížen

Rozdělení výběrových průměrů

- velikost výběrové chyby je dána dvěma charakteristikami:
  - variabilitou znaku v populaci
  - velikostí výběru
- **variabilita znaku v populaci**: čím je vyšší, tím je vyšší i variabilita výběrových průměrů

Rozdělení výběrových průměrů





Rozdělení výběrových průměrů

- velikost výběru – čím větší výběr (*n*), tím méně průměrů výběrů se výrazně odchyluje od průměru populace (= výběrová chyba je menší)

Rozdělení výběrových průměrů

- vzorec pro výpočet výběrové chyby:

$$\sigma_x = \sigma/\sqrt{n}$$

Rozdělení výběrových průměrů

- platí zjednodušení **tzv. centrálního limitního teorému** – pro každou populaci o průměru μ a směrodatné odchylce σ se bude rozdělení výběrových průměrů výběrů (pro rozsah výběru jdoucí do nekonečna) blížit normálnímu rozdělení s průměrem μ a směrodatnou odchylkou $\sigma_x = \sigma/\sqrt{n}$

Rozdělení výběrových průměrů

- **příklad**: když vybereme z populace náhodně vzorek 9 osob, jaká je pravděpodobnost, že jejich průměrné IQ bude větší nebo rovno 112?

Rozdělení výběrových průměrů

- ptáme se vlastně: jaká je pravděpodobnost, že vzorek 9 osob z populace o průměru 100 bude mít průměr 112 nebo vyšší?
- a k tomu potřebujeme znát odpověď na otázku

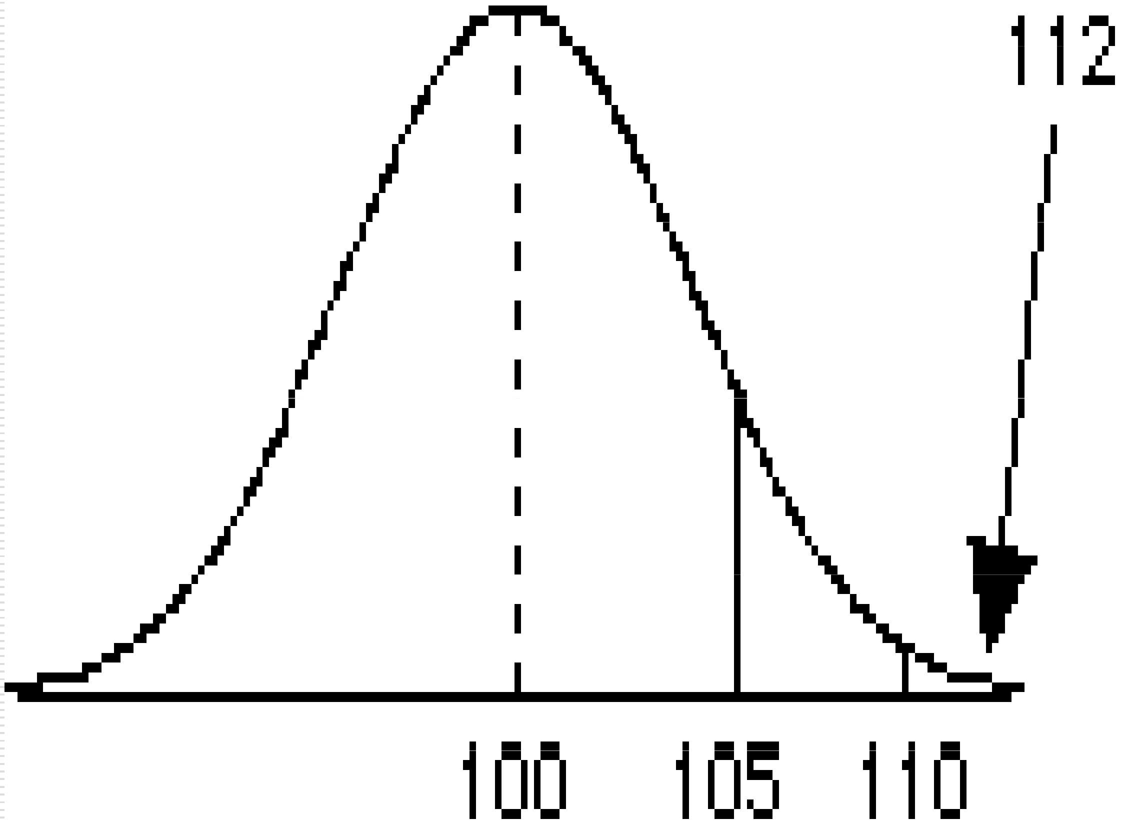
jaké je **rozdělení výběrových průměrů** pro populaci s průměrem 100, sd 15 a velikost vzorku 9?

Rozdělení výběrových průměrů

- musíme zjistit charakteristiku rozdělení výběrových průměrů pro tuto velikost vzorku (N=9) u populace s μ = 100, σ= 15
- průměr RVP = 100
- směrodatná odchylka = výběrová (směrodatná) chyba:

σ_{x} = $\sigma/\sqrt{n}$ = 15/3 = 5

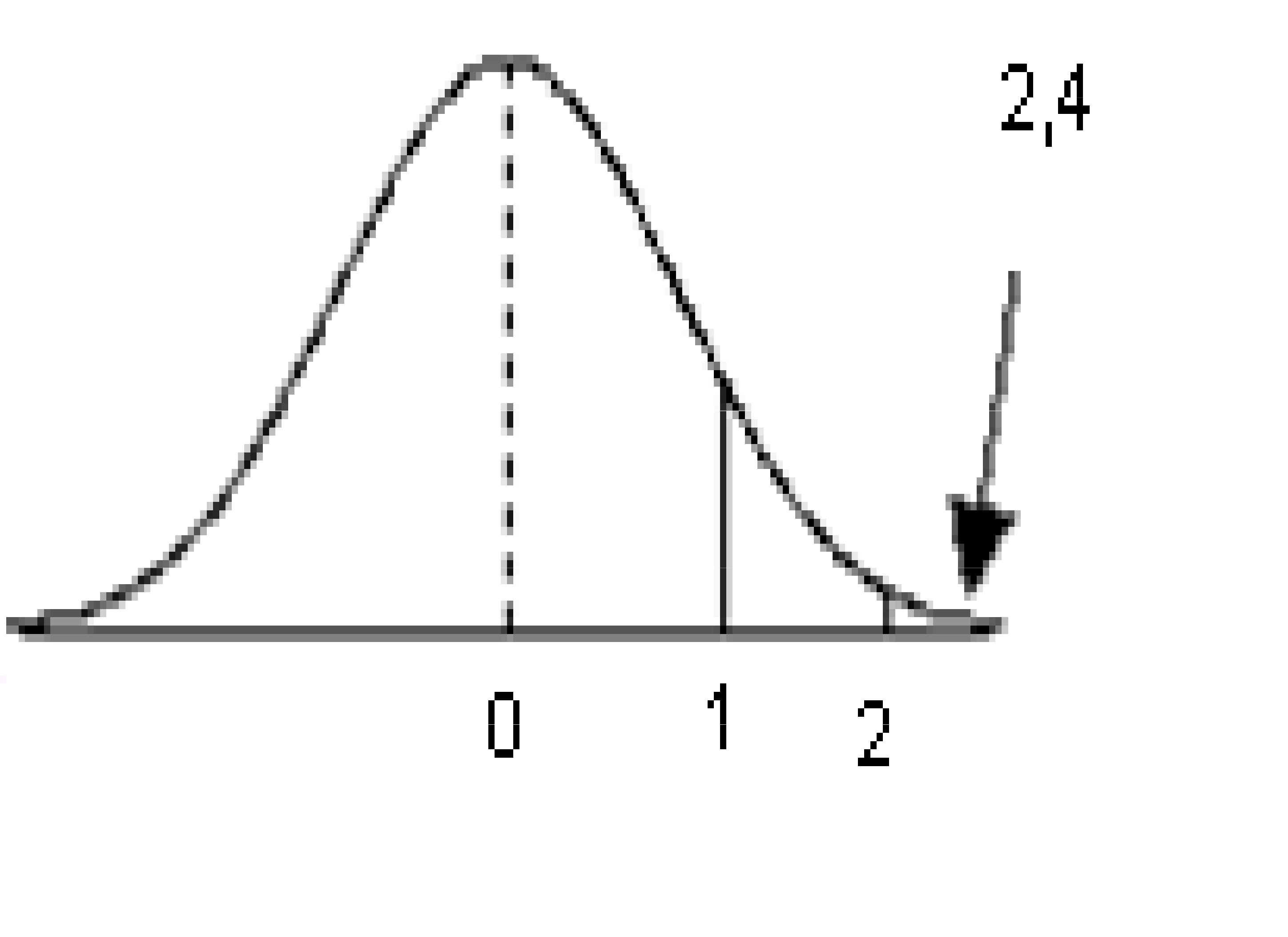
Rozdělení výběrových průměrů



Rozdělení výběrových průměrů

- známe průměr a směrodatnou odchylku rozdělení, převedeme tedy skór 112 na z-skór
- $\mu = 100$, $\sigma_x = 5$
- **z** = $(112-100)/\ \sigma_x = 12/5 =$ **2.4**

Rozdělení výběrových průměrů



Rozdělení výběrových průměrů

- pak najdeme v tabulce z-rozdělení pravděpodobnost pro z=2.4

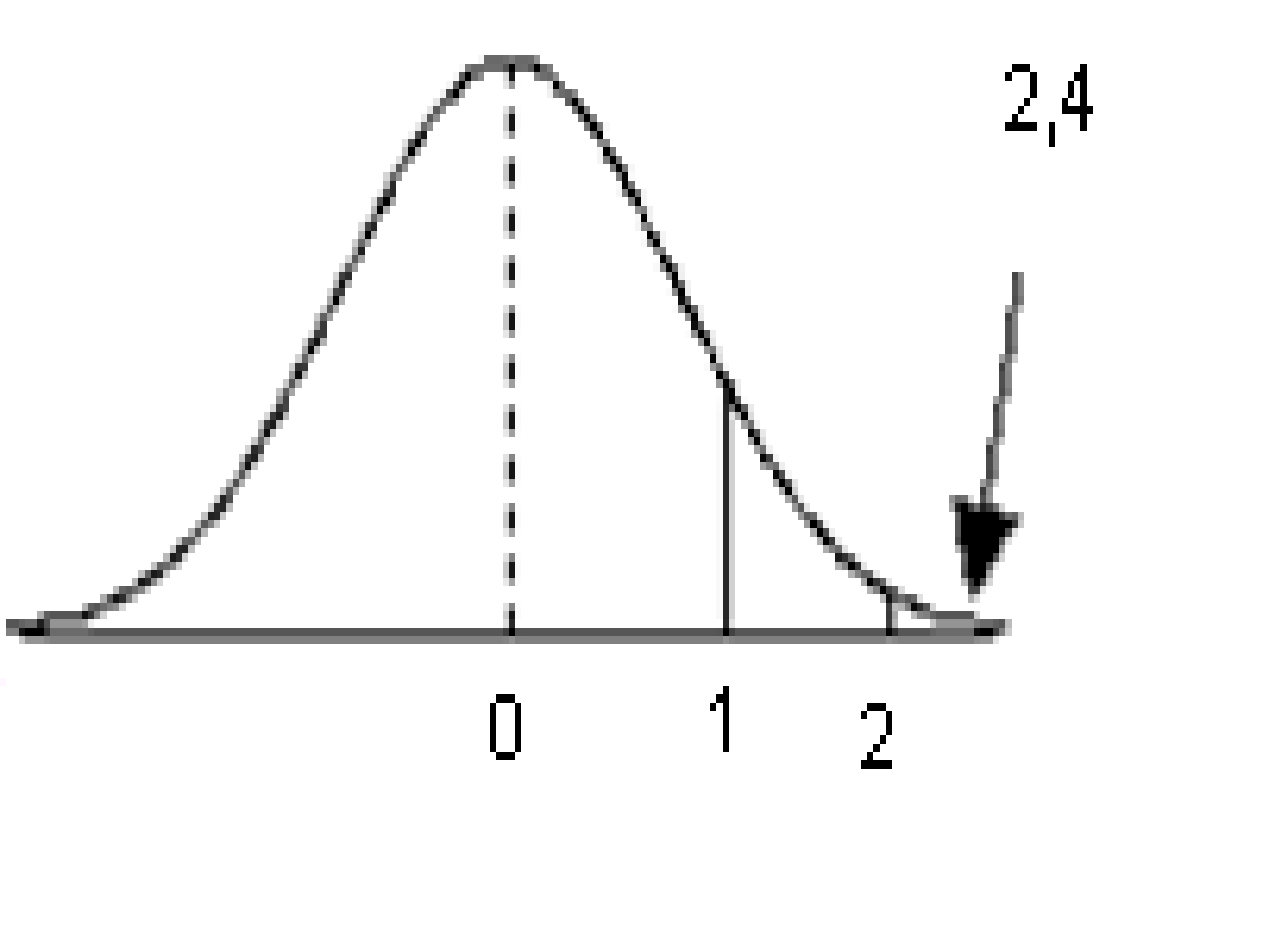
Tabulka z-rozdělení

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
...										
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952

Rozdělení výběrových průměrů

- pak najdeme v tabulce z-rozdělení pravděpodobnost pro z=2.4
- **0.4918**
- přičteme 50% (záporná strana z-rozdělení) = 0.9918
- hodnoty do z=2.4 tvoří 99.18% výběrových průměrů
- zbývá 1-0.9918 = **0.0082**

Rozdělení výběrových průměrů



Rozdělení výběrových průměrů

- **příklad**: když vybereme z populace náhodně vzorek 9 osob, jaká je pravděpodobnost, že jejich průměrné IQ bude větší nebo rovno 112?
- řešení: p = 0,0082 (0,82%)

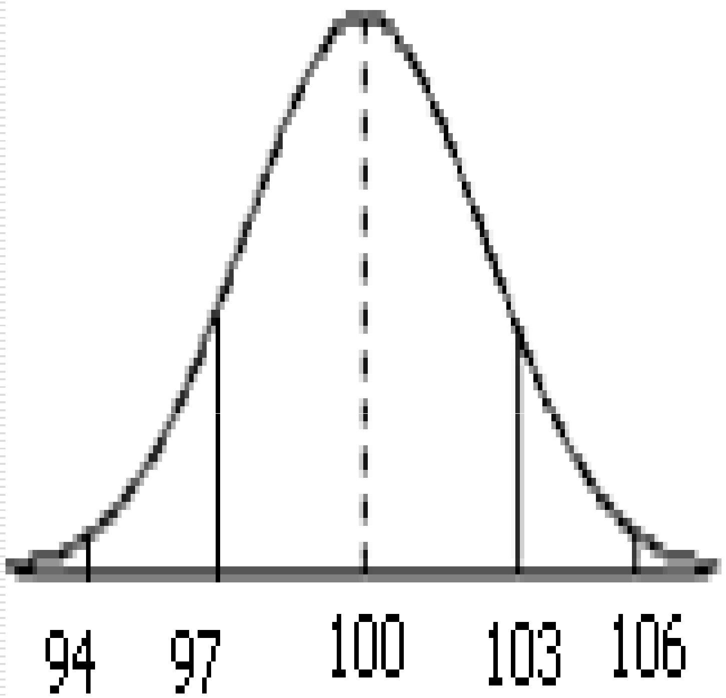
Příklad 2

- IQ (μ=100, σ=15)
- jaká je **pravděpodobnost**, že **průměr výběru** o velikosti n=25 **bude mezi hodnotami** 94 a 106?

Rozdělení výběrových průměrů

- musíme zjistit charakteristiku rozdělení výběrových průměrů pro tuto velikost vzorku (N=25) u populace s μ = 100, σ= 15
- průměr RVP = 100
- směrodatná odchylka = výběrová (směrodatná) chyba:

  $\sigma_x = \sigma/\sqrt{n} = 15/5 = 3$

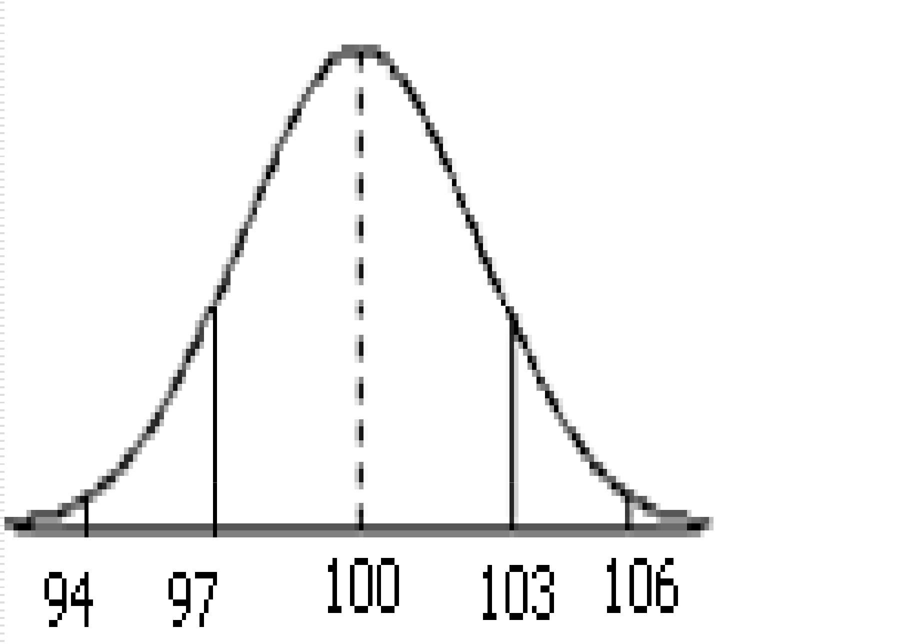


Příklad 2

- IQ (μ=100, σ=15)
- jaká je pravděpodobnost, že průměr výběru o velikosti n=25 bude mezi hodnotami 94 a 106?
- $z = (94-100) / (15/ \sqrt{25}) = -6/3 = -2$
- $z = (106-100) / (15/ \sqrt{25}) = 6/3 = 2$

Příklad 2

- najdeme v tabulce normovaného normálního rozdělení hodnotu pravděpodobnosti pro z=2 a z=-2

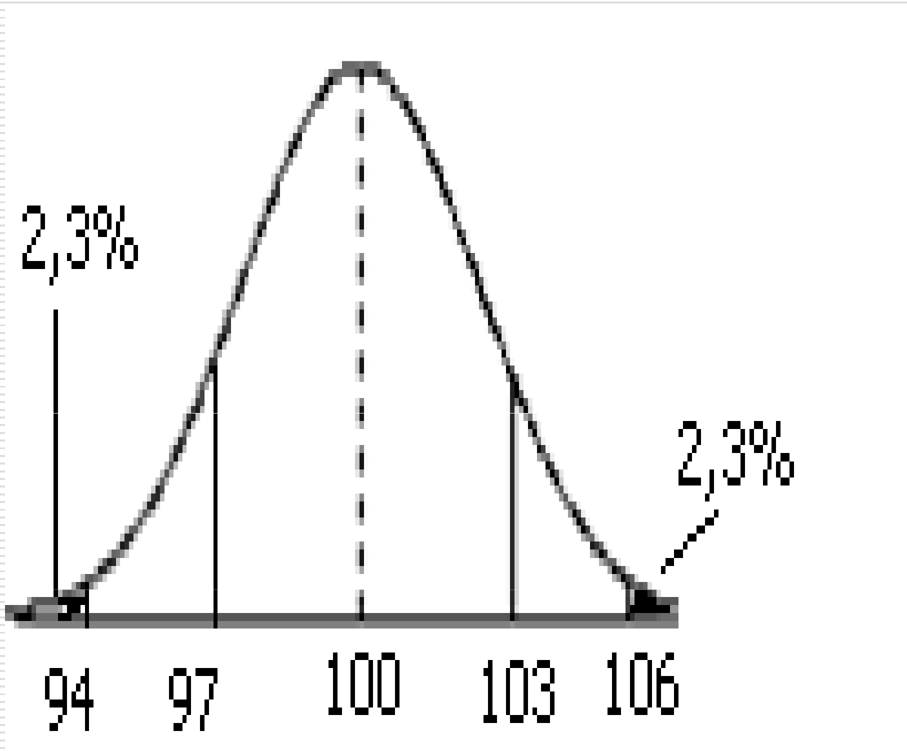


Tabulka z-rozdělení

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
…										
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857

Příklad 2

- hodnota pravděpodobnosti je 0.4772
- sečteme levou a pravou stranu: 0.4772 + 0.4772
- výsledek **0,9544**



Příklad 2

- IQ (μ=100, σ=15)
- jaká je pravděpodobnost, že průměr výběru o velikosti n=25 bude mezi hodnotami 94 a 106?
- **pravděpodobnost takového průměru je 95,4%**

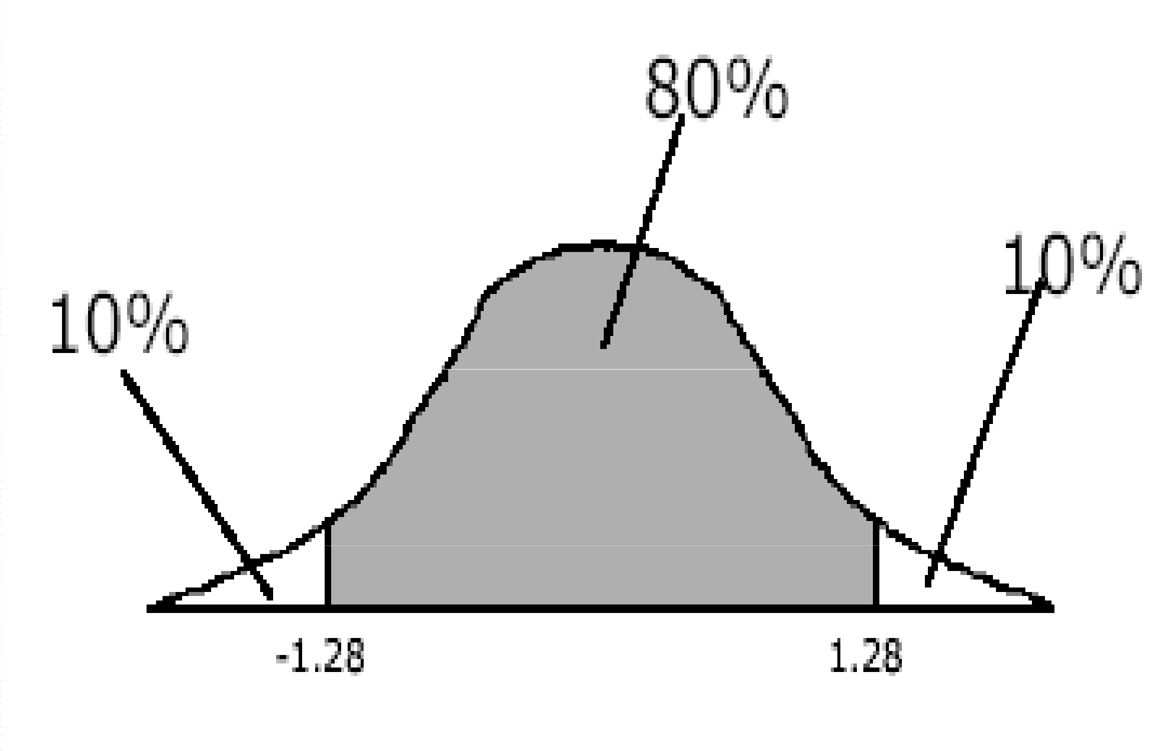
Příklad 3

- IQ (μ=100, σ=15)
- v jakém **rozsahu hodnot** bude pravděpodobně **80% všech průměrů výběrů** o velikosti n=25?

Příklad 3

- IQ (μ=100, σ=15)
- v jakém **rozsahu hodnot** bude pravděpodobně **středních 80% všech průměrů výběrů** o velikosti n=25?
- potřebujeme zjistit hodnotu z, která odděluje pravděpodobnost 10% na obou stranách rozdělení

Příklad 3



Tabulka z-rozdělení

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
...										
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177

Příklad 3

- IQ (μ=100, σ=15)
- v jakém **rozsahu hodnot** bude pravděpodobně **středních 80% všech průměrů výběrů** o velikosti n=25?
- z = 1,28 a -1,28 převedeme na hodnoty IQ

Příklad 3

$$\bar{x} = \mu + z\,(\sigma/\sqrt{n})$$

$$\bar{x} = 100 + 1{,}28\,(15/\sqrt{25}) = 100+1{,}28(3) = \mathbf{103{,}84}$$

$$\bar{x} = 100 + (-1{,}28)\,(15/\sqrt{25}) = 100-1{,}28(3) = \mathbf{96{,}16}$$

Příklad 3

- IQ (μ=100, σ=15)
- v jakém rozsahu hodnot bude pravděpodobně středních 80% všech průměrů výběrů o velikosti n=25?
- **80% výběrových průměrů bude v rozsahu hodnot 96,16 - 103,84**

Kontrolní otázky

- výpočet a především interpretace z-skórů
- normální rozdělení – charakteristiky
- rozdělení výběrových průměrů
- výpočet směrodatné chyby

Literatura

- Hendl: kapitoly 4 a 5