Inferenční statistika - úvod

- z-skóry
- normální rozdělení
- pravděpodobnost
- rozdělení výběrových průměrů

Pravděpodobnost

- postupy induktivní statistiky vycházejí z teorie pravděpodobnosti
- **pravděpodobnost**, že nastane určitý výsledek, **definujeme** jako podíl

$$P\,(A) = \frac{\text{počet pokusů, kdy nastal jev A}}{\text{celkový počet jevů}}$$

Pravděpodobnost

- pravděpodobnost bývá uváděna nejčastěji jako **podíl** (0,33), **zlomek** (1/3) nebo **procento** (33,3%)
- pravděpodobnost určitého jevu nebo třídy jevů můžeme odhadnout z rozdělení hodnot (četností)

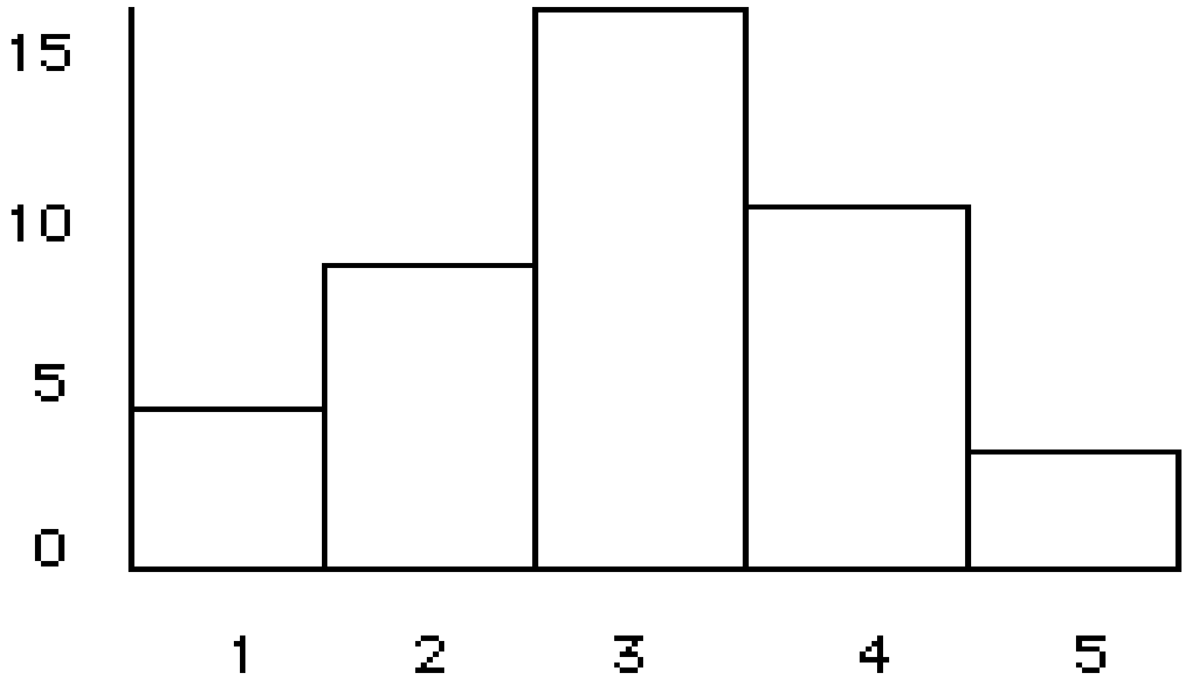
Pravděpodobnost - příklady

- představme si, že máme krabici se 40 očíslovanými žetony s čísly 1 – 5
- v tabulce jsou uvedeny absolutní i relativní četnosti jednotlivých čísel žetonů

Pravděpodobnost

X	*f*	*p*
1	**4**	0,10
2	**8**	0,20
3	**16**	0,40
4	**10**	0,25
5	**2**	0,05

Pravděpodobnost



Pravděpodobnost - příklady

- vaším úkolem je vytáhnout 1 žeton
- **jaká je pravděpodobnost, že vytáhnete žeton s číslem 3?**

Pravděpodobnost

X	*f*	*p*
1	**4**	0,10
2	**8**	0,20
3	**16**	0,40
4	**10**	0,25
5	**2**	0,05

Pravděpodobnost

- vaším úkolem je vytáhnout 1 žeton
- jaká je pravděpodobnost, že vytáhnete žeton s číslem 3?
- **p (3) = f/N = 16/40 =0,40**
  nebo 2/5 či 40%

Pravděpodobnost

- **Jaká je pravděpodobnost, že vytáhnete žeton s číslem vyšším než 2?**

Pravděpodobnost

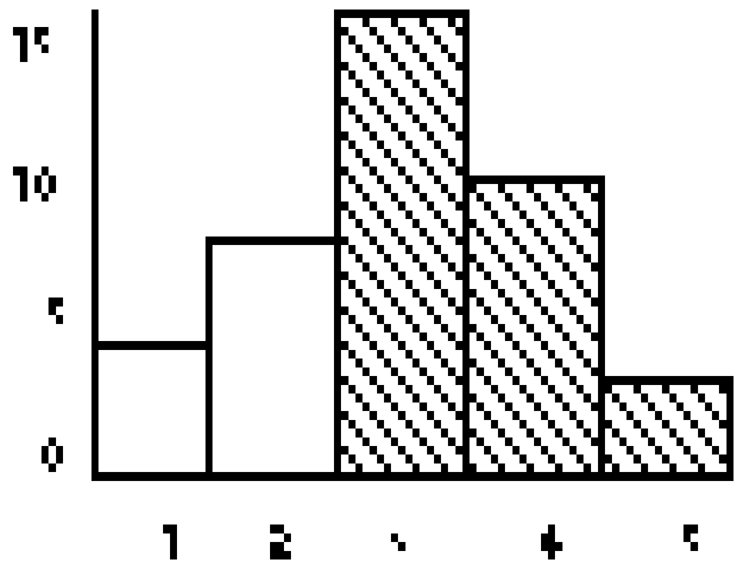
X	*f*	*p*
1	**4**	0,10
2	**8**	0,20
3	**16**	0,40
4	**10**	0,25
5	**2**	0,05

Pravděpodobnost

- Jaká je pravděpodobnost, že vytáhnete žeton s číslem vyšším než 2?

  $p(X > 2) = ?$

  0,05 + 0,25 + 0,40 = **0,70**



Pravděpodobnost

- **Jaká je pravděpodobnost, že vytáhnete žeton s číslem nižším než 5?**

Pravděpodobnost

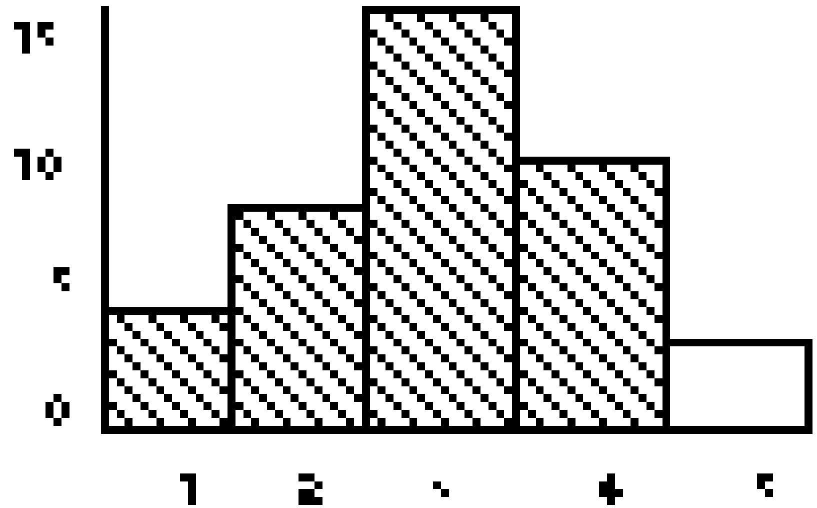
X	*f*	*p*
1	**4**	0,10
2	**8**	0,20
3	**16**	0,40
4	**10**	0,25
5	**2**	0,05

Pravděpodobnost

- Jaká je pravděpodobnost, že vytáhnete žeton s číslem nižším než 5?

  $p(X < 5) = ?$

  0,10 + 0,20 + 0,40 + 0,25 = **0,95**



Pravděpodobnost

- **Jaká je pravděpodobnost, že vytáhnete žeton s číslem nižším než 4 a vyšším než 1?**

Pravděpodobnost

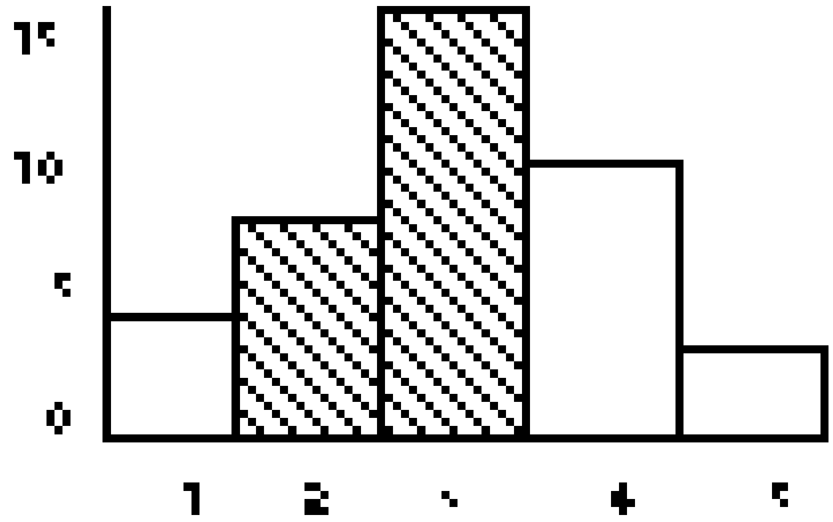
X	*f*	*p*
1	**4**	0,10
2	**8**	0,20
3	**16**	0,40
4	**10**	0,25
5	**2**	0,05

Pravděpodobnost

- Jaká je pravděpodobnost, že vytáhnete žeton s číslem nižším než 4 a vyšším než 1?

  $p(4 > X > 1) = ?$

  0,20 + 0,40 = **0,60**



Pravděpodobnost

- pravděpodobnost odpovídá hustotě **oblasti pod křivkou** pro daný interval

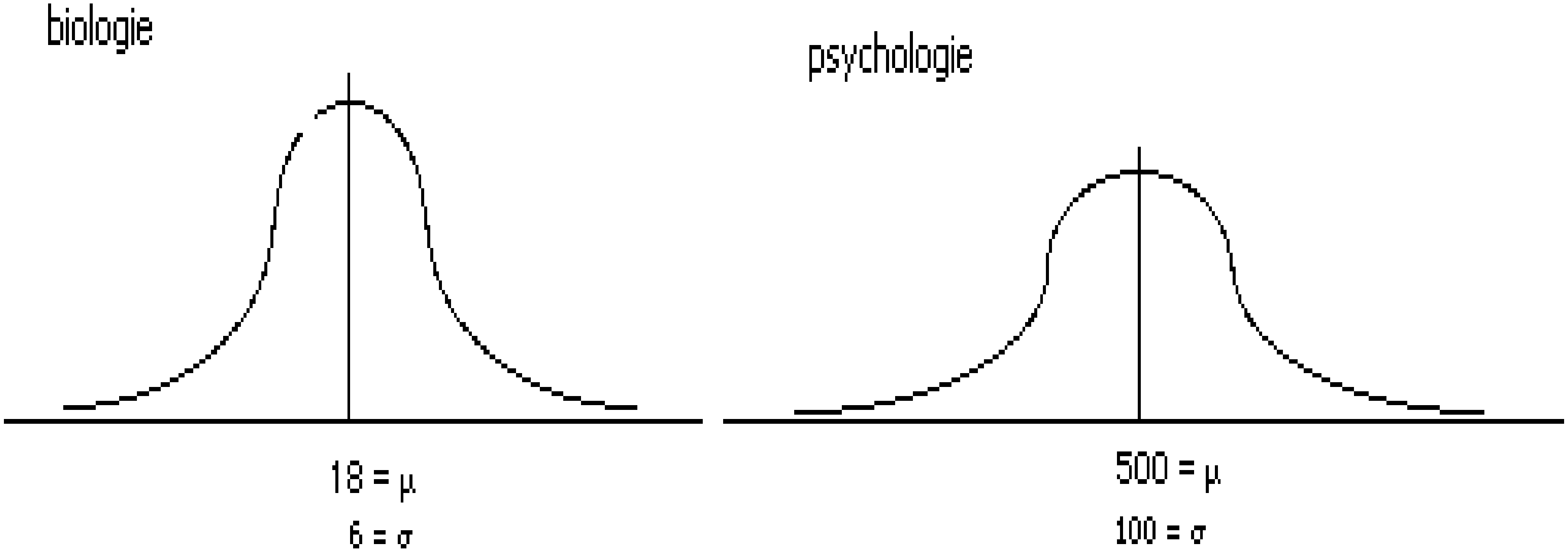
Z-skóry

- transformace hodnot proměnné
- umožňují najít a popsat **pozici každé hodnoty** v rámci rozdělení hodnot
- a také **srovnávání hodnot** pocházejících z měření **na rozdílných stupnicích**
- hrubé skóry jsou převedeny na **standardizovanou stupnici** (jednotkou je směrodatná odchylka)

Z-skóry - příklad

- např. skóry ze dvou testů – biologie a psychologie
- v testu z biologie byl průměr celého ročníku m=18 (sd=6)
- v testu z psychologie byl průměr celého ročníku m=500 (sd=100)
- student získal 26 bodů z biologie a 620 z psychologie. Ve kterém předmětu byl lepší?

Z-skóry - příklad



Z-skóry

- přímé porovnání není snadné – skóry z obou testů mají rozdílné průměry i směrodatné odchylky
- z skór =odchylka skóru od průměru vzhledem k velikosti směrodatné odchylky
- z = odch. od průměru/směr. odch.

Z-skóry - příklad

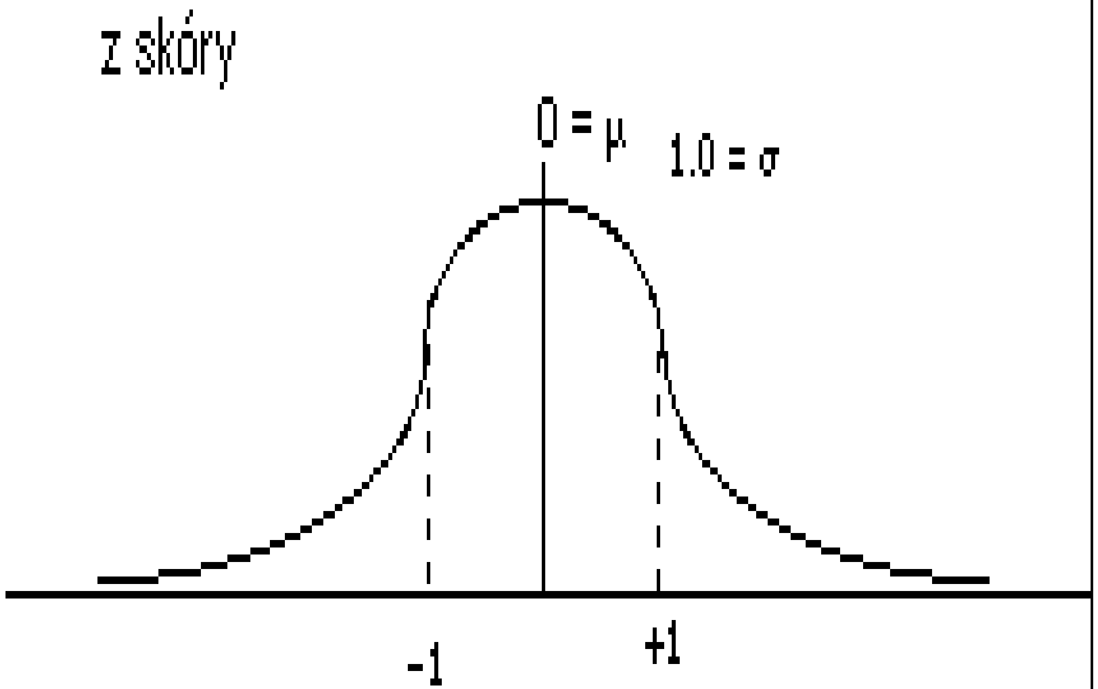
- skór z biologie: (26-18)/6 = 1,33
- skór psychologie: (620-500)/100=1,2
- v biologii byl student lepší: 1,33 směrodatné odchylky nad průměrem

Z-skóry

- z-skór přesně udává pozici každé hodnoty vzhledem k ostatním hodnotám
- znaménko (+ nebo -) ukazuje, zda je hodnota nad nebo pod průměrem rozdělení
- hodnota z-skóru upřesňuje, kolik směrodatných odchylek byla hodnota od průměru vzdálena

Z-skóry

- průměr rozdělení z-skórů je vždy 0
- směrodatná odchylka je 1



Z-skóry

vzorec pro výpočet z-skóru hodnoty X

- u populace: $z = (X_i - \mu) / \sigma$

- u vzorku: $z = (X_i - m) / s$

Z-skóry

- podobně můžeme i z-skór převést na hrubý skór, známe-li průměr a směrodatnou odchylku

Z-skóry

- např. u stupnice IQ
- $\mu = 100, \sigma = 15$
- pro osobu se z=-3 (3 směrodatné odchylky pod průměrem) bude IQ ?

Z-skóry

- např. u stupnice IQ $\mu = 100$, $\sigma = 15$
- pro osobu se z=-3 (3 směrodatné odchylky pod průměrem) bude IQ

$$X = Z \cdot \sigma + \mu$$

$$X = -3 \cdot 15 + 100$$

$$X = 55$$

Rozdělení z-skórů

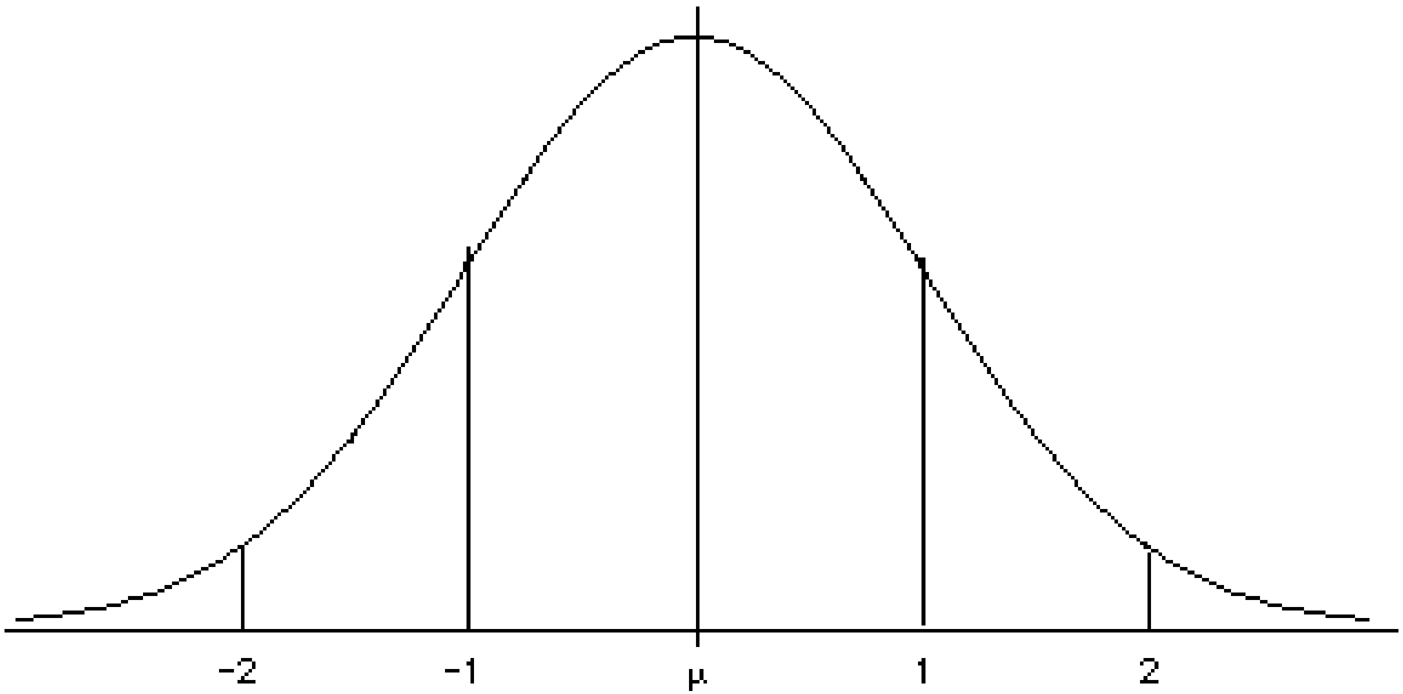
- **tvar** rozdělení z-skórů je **stejný** jako tvar původního rozdělení hrubých skórů
- průměr je 0, směrodatná odchylka 1
- transformace změní jen označení hodnot na ose X

Normální rozdělení

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

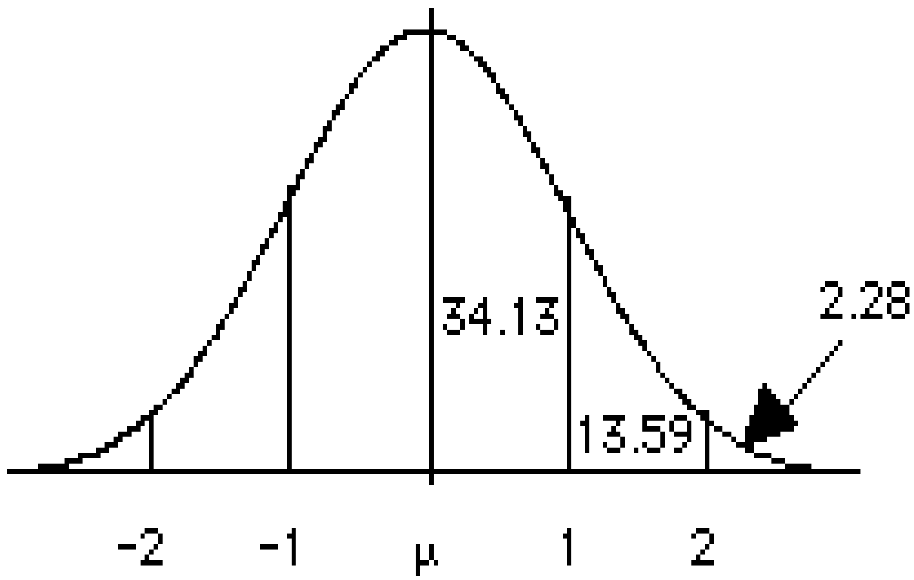
$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2}$$

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 mezi 2. a 3. směr. odchylku

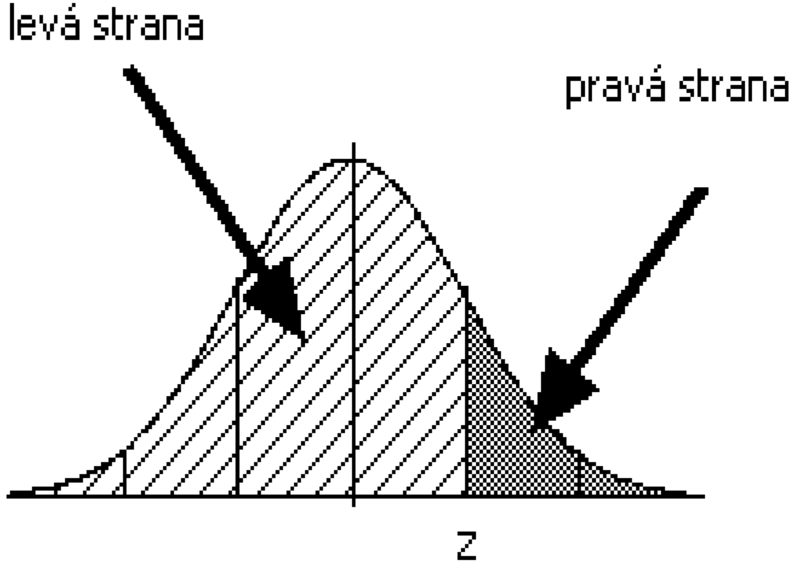


Normální rozdělení

- 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

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621

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

- Kolik procent osob z populace má IQ 130 nebo vyšší? (μ = 100, σ =15)

otázku je možno formulovat i jako:

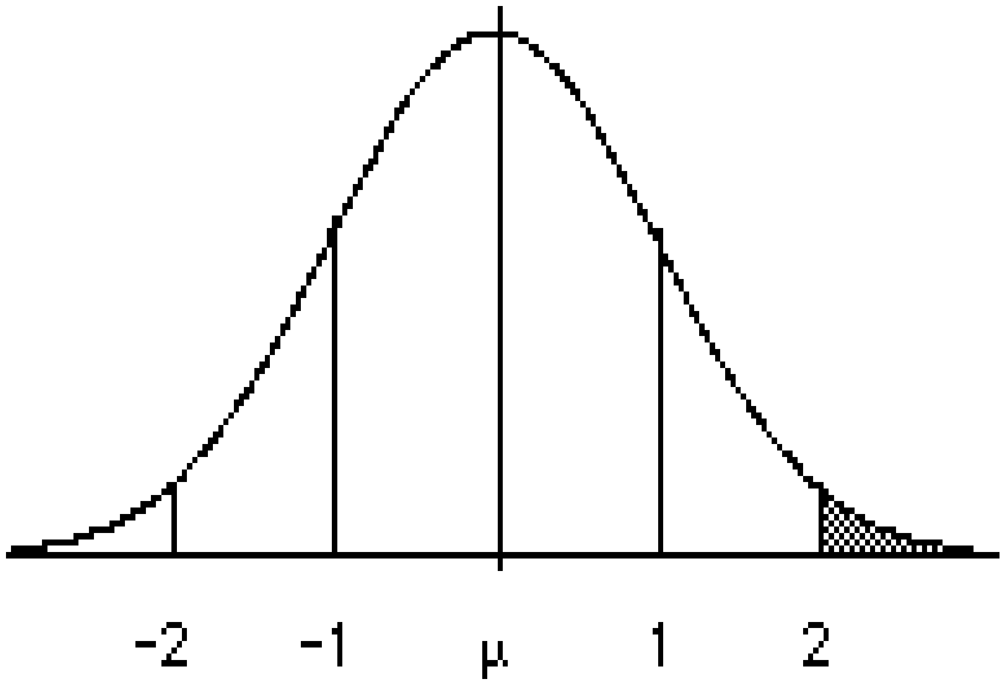
- Jaká je pravděpodobnost, že náhodně vybraná osoba z populace bude mít IQ 130 nebo vyšší? (μ = 100, σ =15)

Normální rozdělení - příklady

- Jaká je pravděpodobnost, že náhodně vybraná osoba z populace bude mít IQ 130 nebo vyšší? (μ = 100, σ =15)
- z = (130 – 100)/15
- **z = 2**

Normální rozdělení - příklady

- z = 2



Normální rozdělení

1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
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
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890

Normální rozdělení - příklady

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

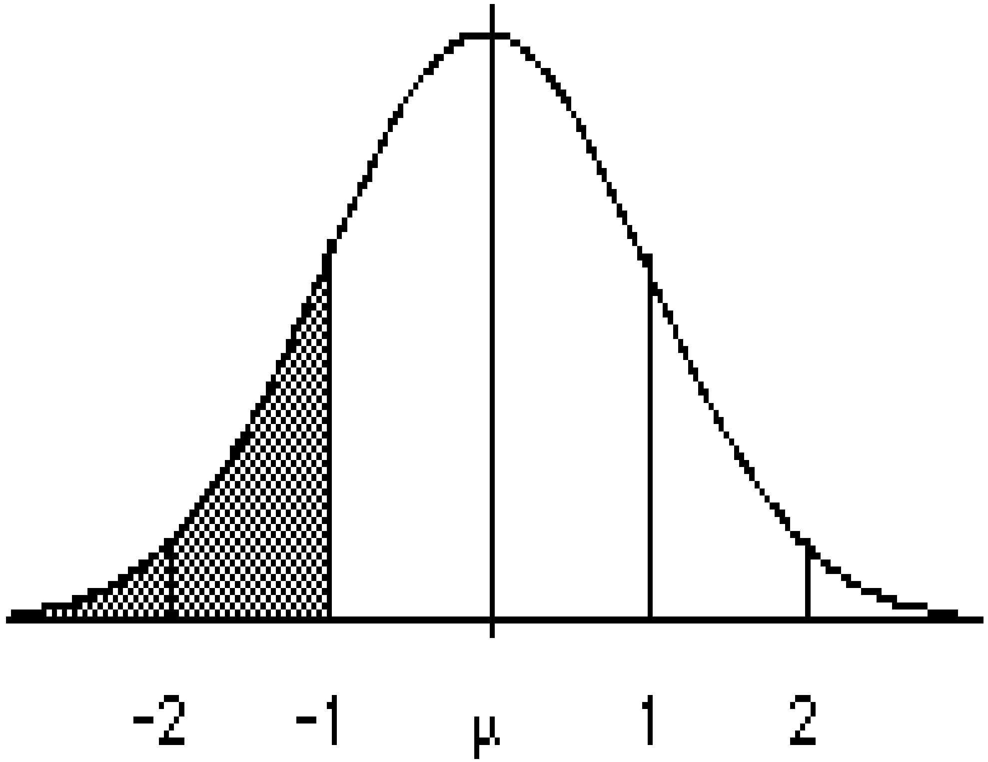
Normální rozdělení - příklady

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

Normální rozdělení - příklady

z = (100-85)/15

z = -1



Normální rozdělení

0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621
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

Normální rozdělení - příklady

- Jaká je pravděpodobnost, že náhodně vybraná osoba z populace bude mít IQ 85 nebo nižší?
- $z = -1$
- $p = 0.5 – 0.3413 =$ **0.1587**
- tj. **15,9%**

Normální rozdělení - příklady

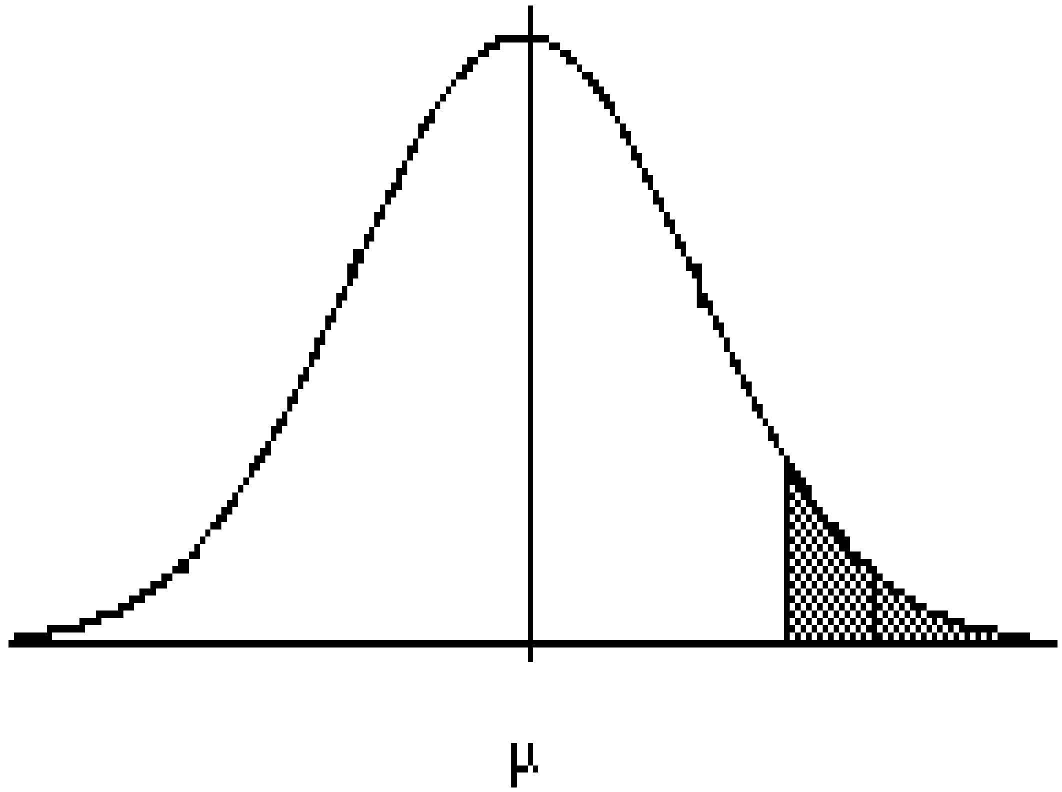
- postup při zjišťování **z-skóru** z tabulky:
  - načrtnout si normální rozdělení
  - vystínovat oblast odpovídající zadané pravděpodobnosti
  - v tabulce vyhledat příslušný z-skór
  - vypočítat z něj hrubý skór

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

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
...										
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633

Normální rozdělení - příklady

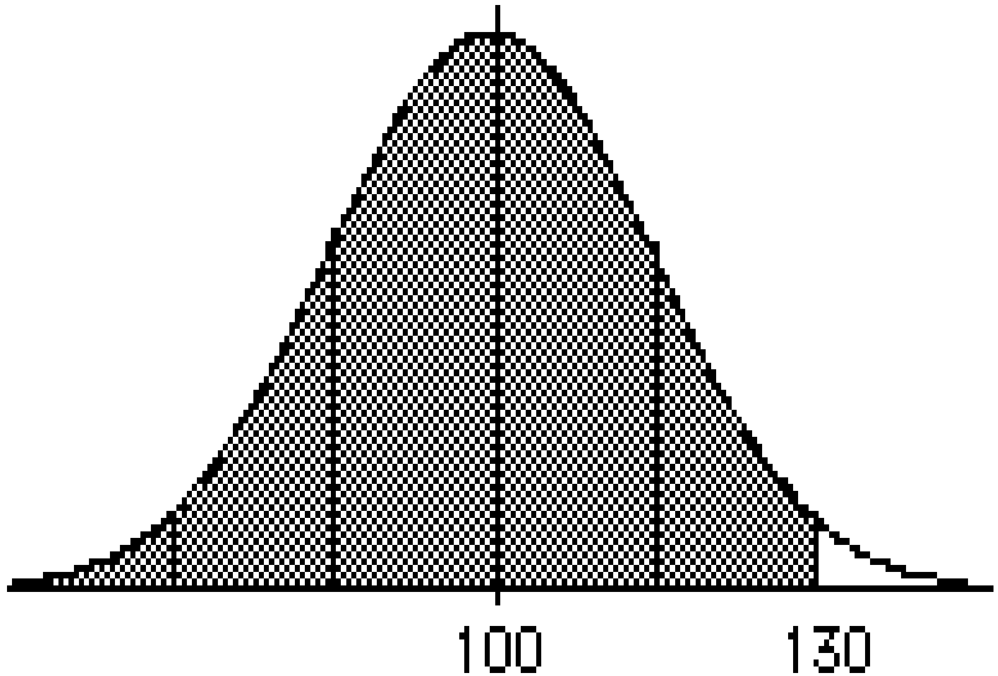
- Jakou minimální hodnotu IQ musí člověk mít, aby patřil mezi 5% osob s nejvyššími hodnotami IQ?
- p = 0.05
- z tabulky: **z = 1.645**
- X = (1.645)*(15) + 100 = **124.675**
- musí mít IQ 124 bodů

Normální rozdělení - příklady

- pomocí tabulky normálního rozdělení je možno nalézt také hodnotu percentilu
- příklad: kolik procent osob má nižší hodnoty IQ než člověk s IQ 130?
- $z = +2$

Normální rozdělení - příklady

- z = 2



Normální rozdělení

1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
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
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890

Normální rozdělení - příklady

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

p = 0.4772 (+ 50% pod průměrem)

97.72% osob má nižší skór

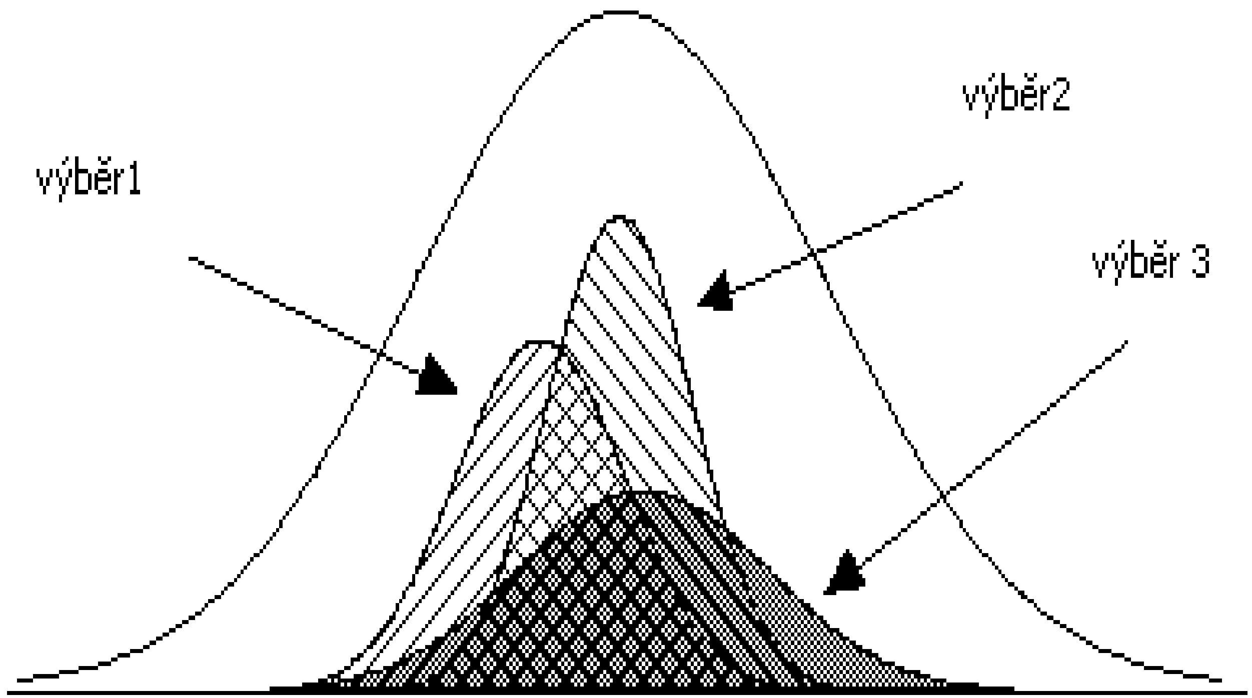
Rozdělení výběrových průměrů

- cílem induktivní statistiky je odhadnout <u>parametry populace</u> z <u>charakteristik vzorku</u> (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)

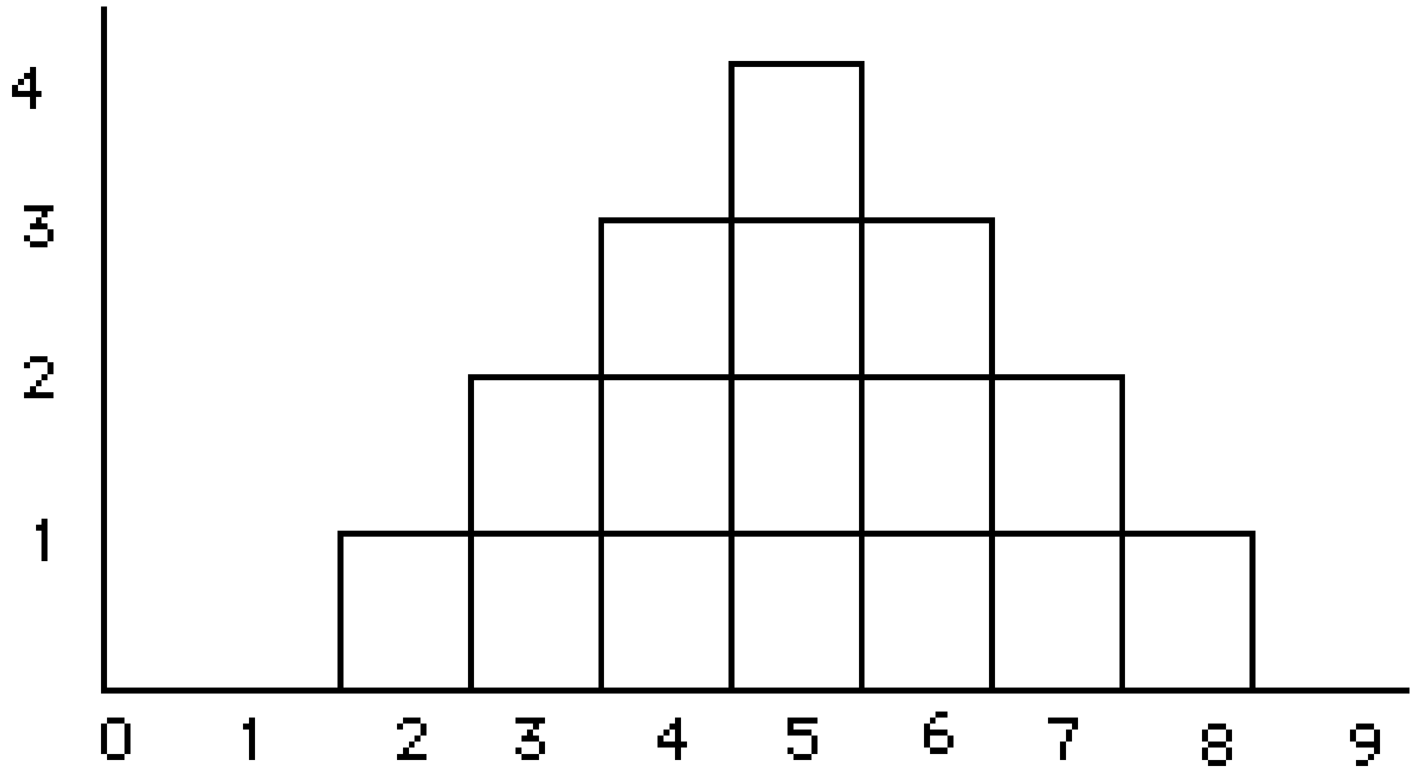
Rozdělení výběrových průměrů

- **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

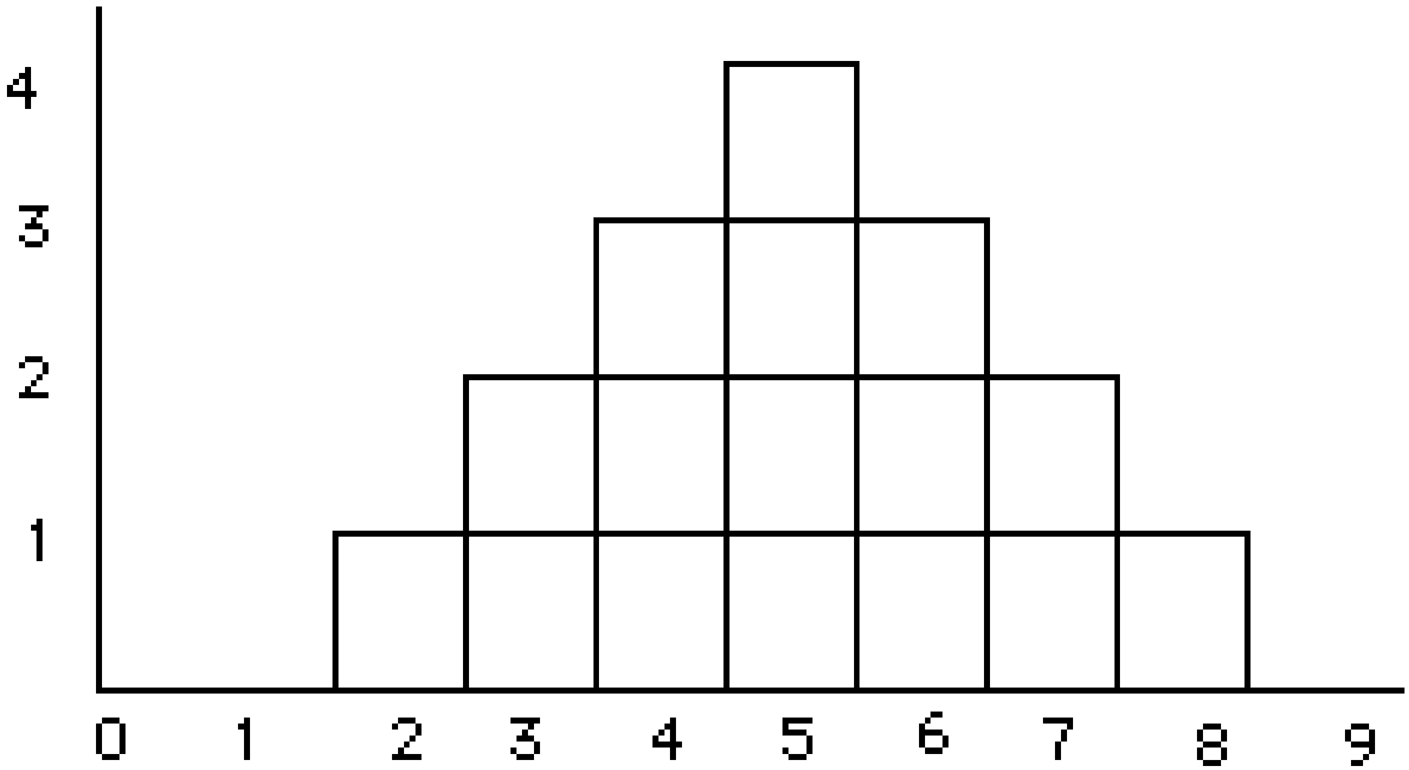
Rozdělení výběrových průměrů



Rozdělení výběrových průměrů

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

Rozdělení výběrových průměrů



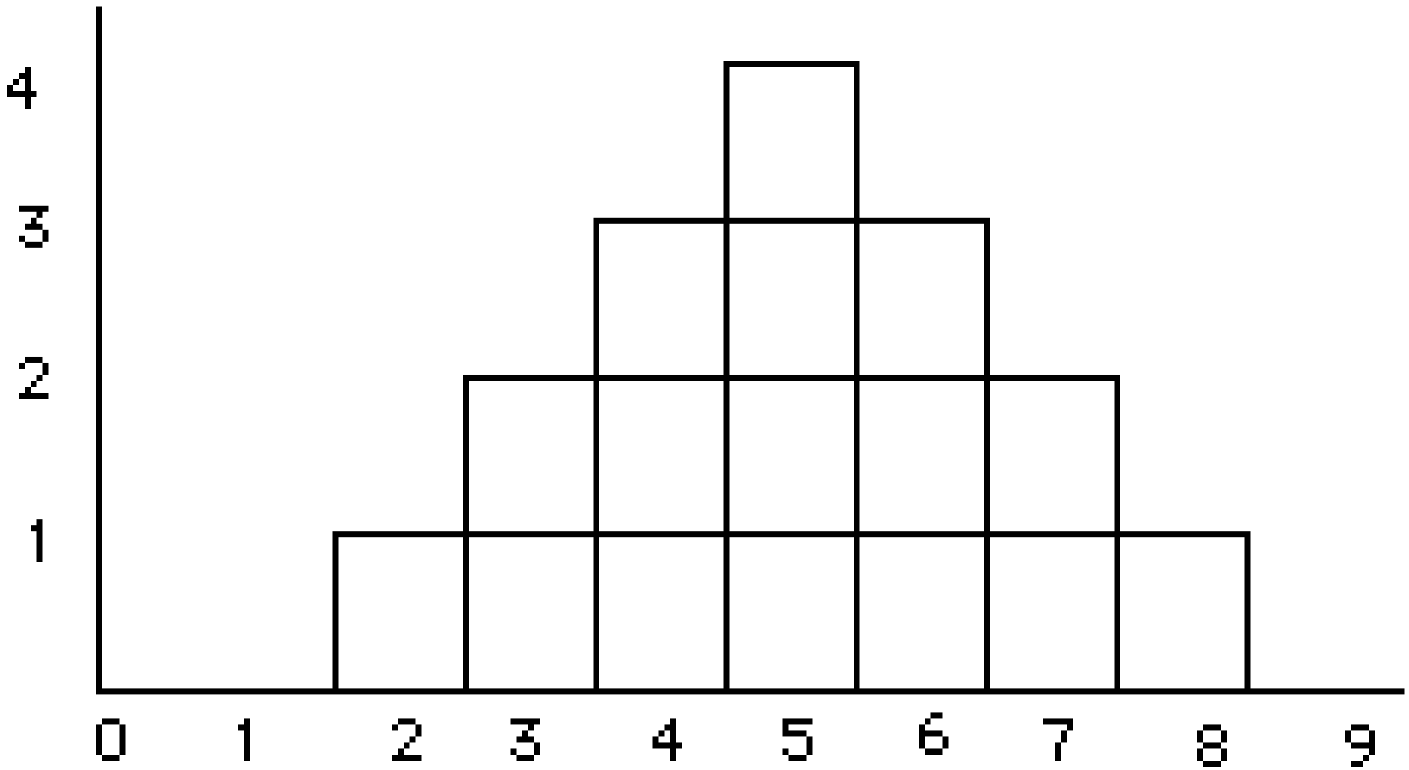
Rozdělení výběrových průměrů

- 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 a více) 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

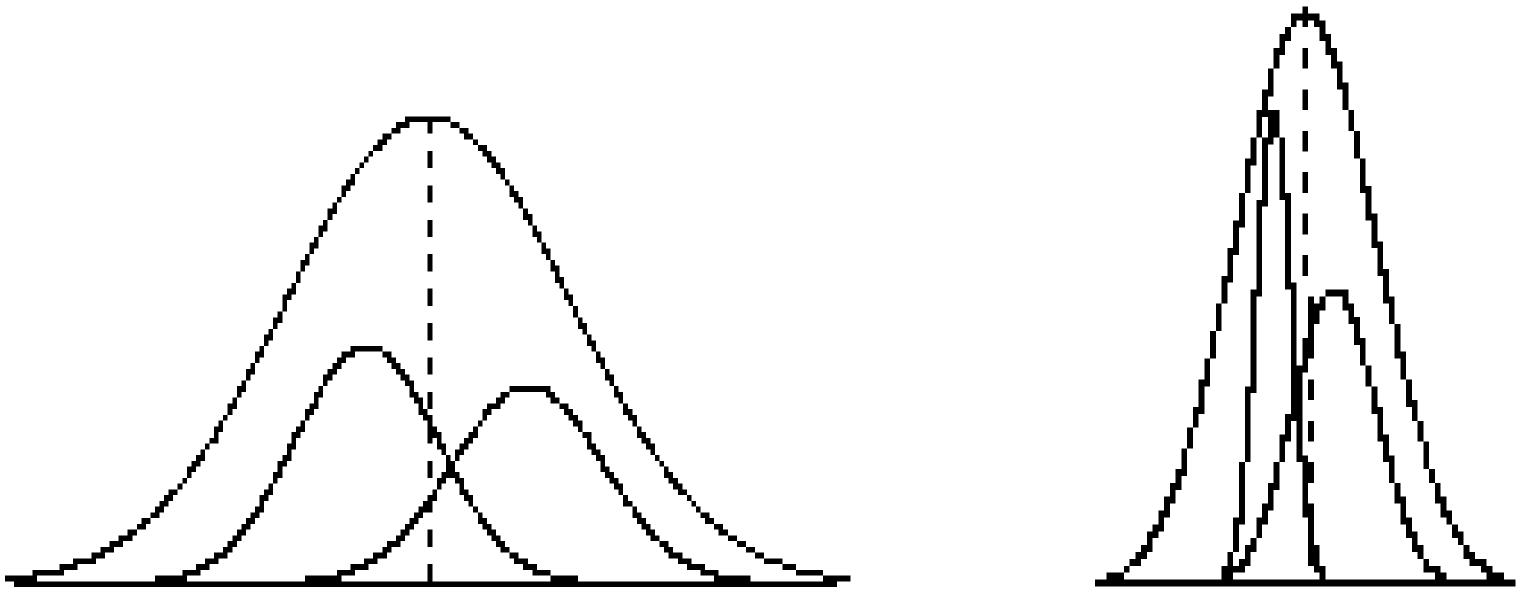
Rozdělení výběrových průměrů

- **variabilita** – směrodatná odchylka RVP se označuje jako výběrová nebo standardní chyba průměru (standard error)
- jde o směrodatnou odchylku výběrových průměrů od průměru populace
- 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 v populaci a 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 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

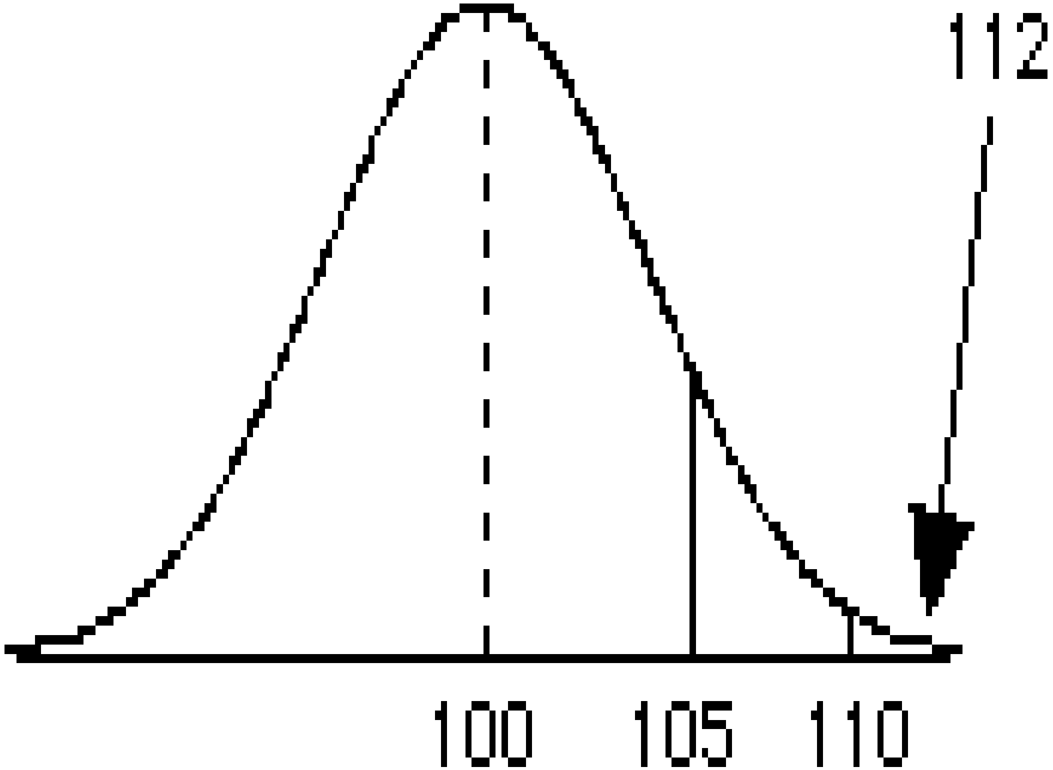
- ptáme se vlastně: jaká je pravděpodobnost, že vzorek 9 osob z populace o průměru 100 bude mít průměr 112 nebo vyšší?

Rozdělení výběrových průměrů

- musíme zjistit charakteristiku rozdělení výběrových průměrů pro tuto velikost vzorku (N=9) u populace s $\mu = 100$, $\sigma = 15$
- průměr RVP = 100
- směrodatná odchylka = standardní chyba:

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

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

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

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
- z tabulky $P(Z \geq 2.4)$ = **0.4918**
- odečteme od 50% (celá jedna strana z-rozdělení) a vyjde nám pravděpodobnost:
- p = 0.5000 – 0.4918 = **0.0082**

Kontrolní otázky

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

Literatura

- Hendl: kapitoly 4 a 5