



Analýza dát

Prednáška 4

Náhodná premenná

Normálne rozdelenie

Poznámka o populácii a vzorke

- **Populácia**, ináč **základný súbor**, je množina všetkých možných prvkov, respondentov, vyhovujúca fenoménu nášho záujmu.
- **Vzorka** alebo tiež **výberový súbor** je množina štatistických jednotiek, ktoré boli z populácie vybrané podľa vopred stanovených pravidiel.
- **Náhodný výber** = taký výber štatistických jednotiek z populácie, kde každá štatistická jednotka má rovnakú (nenulovú) pravdepodobnosť výberu do vzorky.
- **Názvoslovie pre populáciu a pre vzorku:**
 - **Aritmetický priemer**
 - **Smerodajná odchýlka**
 - **Rozptyl**

Náhodná premenná

- **Náhodná premenná X** je funkcia, ktorá priraduje každému prvku (náhodnému javu) numerickú hodnotu (pravdepodobnosť náhodného javu).
- Príklady náhodných premenných:
 - Počet chlapcov vo štvorici súrodencov
 - Súčet bodov pri hode dvomi kockami
 - Zisk z jedného študenta pri kúpe učebníc
 - ...

Typ náhodnej premennej

- **Diskrétna náhodná premenná:** má konečný alebo spočítateľný počet možných hodnôt;
- **Spojité náhodná premenná:** počet jej možných hodnôt je veľmi veľký, často závislý od presnosti merania (váha, dĺžka, čas, rýchlosť, cena)

Pravidlá rozdelenia pravdepodobnosti náhodnej premennej

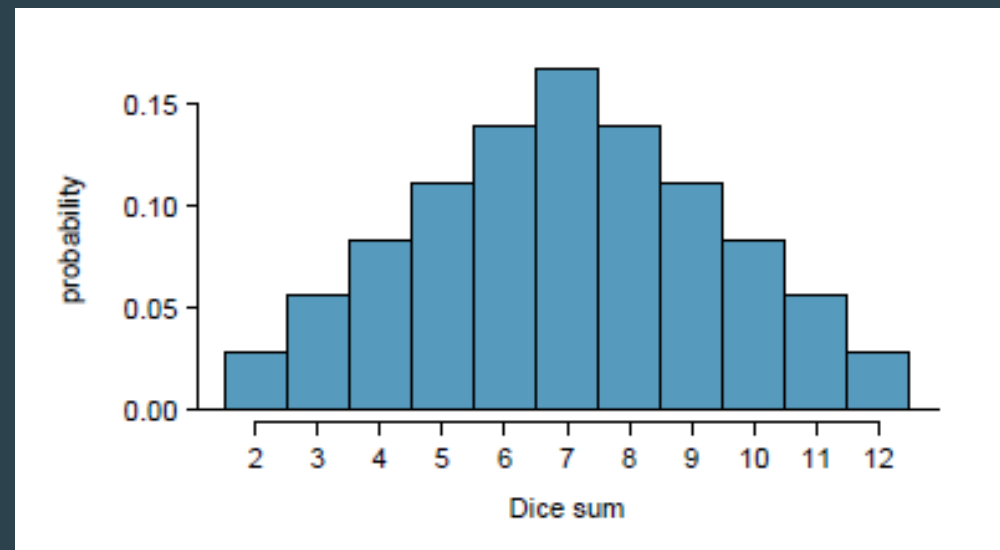
- Všetky hodnoty náhodnej premennej musia predstavovať disjunktné náhodné javy
- Každá pravdepodobnosť je číslo z intervalu $\langle 0,1 \rangle$
- Súčet všetkých pravdepodobností je 1

Diskrétna náhodná premenná

Rozdelenie pravdepodobnosti diskkrétnej náhodnej premennej

- Je tabuľka, ktorá obsahuje všetky možné hodnoty náhodnej premennej a im priradené pravdepodobnosti
- Vezmime náhodnú premennú X =súčet bodov, ktoré padnú pri hode dvoma kockami a vytvorme tabuľku rozdelenia jej pravdepodobností

X	P(X)
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36



Stredná hodnota diskkrétnej náhodnej premennej

- Je miera centrality rozdelenia pravdepodobnosti. Je to vážený priemer, kde hodnoty náhodnej premennej sú „vážené“ ich pravdepodobnosťami.
- Strednú hodnotu tiež niekedy nazývame očakávanou hodnotou, pretože je to hodnota, ktorej nastatie priemerne očakávame

Stredná hodnota diskkrétnej náhodnej premennej

- Nech $x_1, x_2, \dots, x_n$ sú všetky možné hodnoty diskkrétnej náhodnej premennej X , ktorých pravdepodobnosti sú $P(x_1), P(x_2), \dots, P(x_n)$. Stredná hodnota (očakávaná) tejto náhodnej premennej je daná vzťahom:
- $\mu = E(X) = \sum_{i=1}^n x_i P(x_i)$

Rozptyl a smerodajná odchýlka diskkrétnej náhodnej premennej

- Nech $x_1, x_2, \dots, x_n$ sú všetky možné hodnoty diskkrétnej náhodnej premennej X , ktorých pravdepodobnosti sú $P(x_1), P(x_2), \dots, P(x_n)$ a nech stredná hodnota náhodnej premennej je $\mu_X = E(X)$. Potom rozptyl $V(X)$ a smerodajná odchýlka $SD(X)$ náhodnej premennej X sú dané vzťahmi:
- $V(X) = \sigma_X^2 = \sum_{i=1}^n (x_i - \mu_X)^2 P(x_i)$
- $SD(X) = \sigma_X = \sqrt{V(X)} = \sqrt{\sum_{i=1}^n (x_i - \mu_X)^2 P(x_i)}$

Príklad

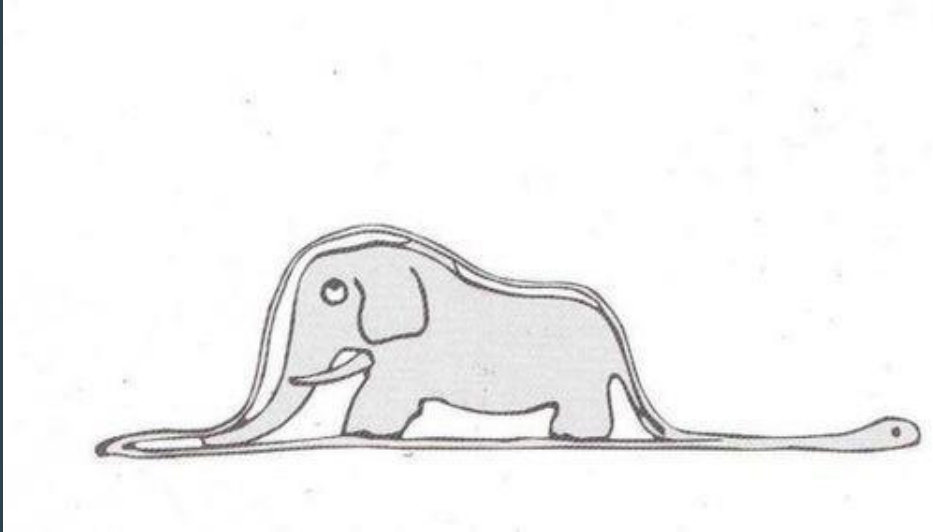
Na predmet AD (na nemenovanej univerzite) je možné si kúpiť učebnicu za 137 € a zbierku úloh za 33€. Z dlhodobých záznamov o kúpe kníh je zrejmé, že 55% študentov v príslušnom akademickom roku si kúpi iba učebnicu, 25% študentov aj učebnicu aj zbierku a 20% sa zaobíde bez akejkoľvek literatúry. (pozn.: nikto si nekúpil iba zbierku úloh).

- čo môžeme považovať v tomto prípade za náhodnú premennú?
- aký bude priemerný zisk na jedného študenta, t.j. $E(X)$?
- aký bude rozptyl $V(X)$ a smerodajná odchýlka $SD(X)$ náhodnej premennej?

Príklad – riešenie

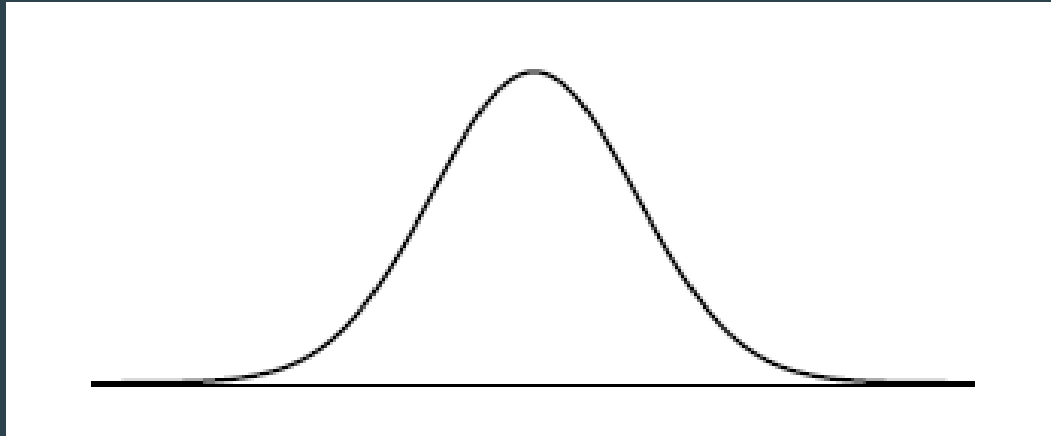
Spojitá náhodná premenná

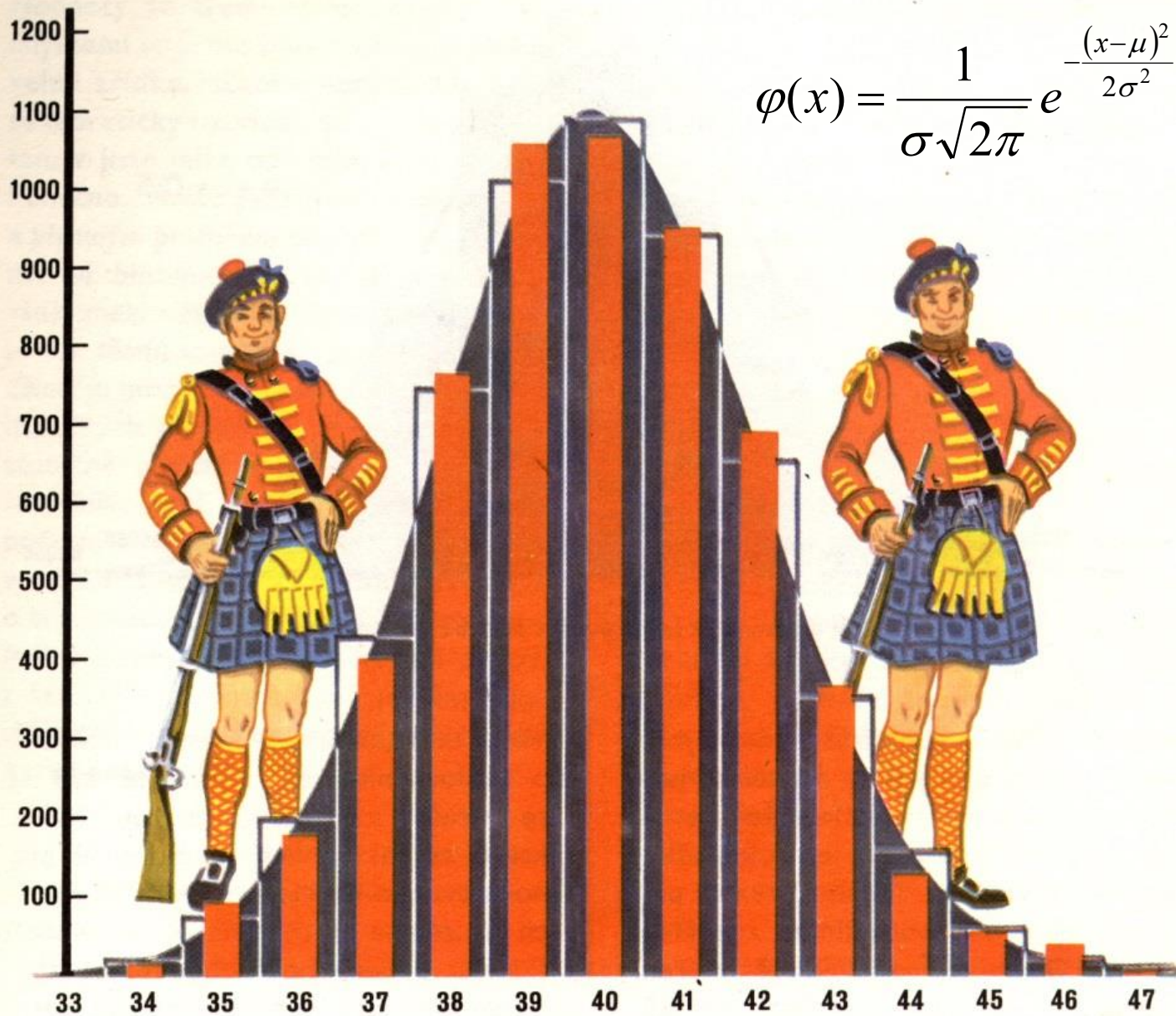
Normálne rozdelenie



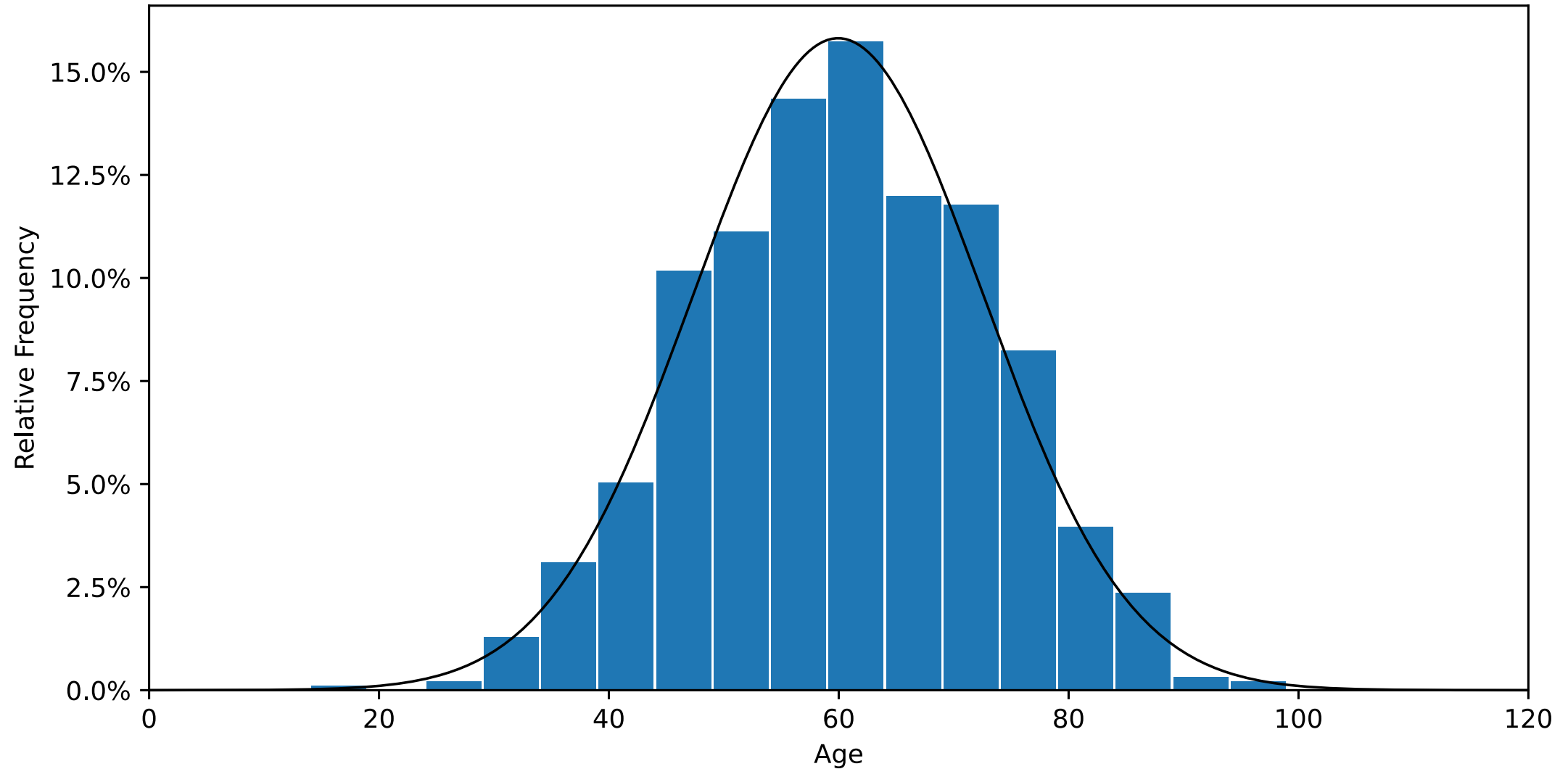
Normálne rozdelenie

- Najznámejšie a najčastejšie používané rozdelenie **spojitej** náhodnej premennej:
- Príklady: výška jedincov, IQ jedincov, hmotnosť detí pri narodení, mesačný príjem jedincov v krajine, veľkosť nohy, skóre na skúške zo štatistiky...
- Histogram má tvar gaussovej krivky = bell shaped (krivka tvaru zvona):
 - Symetrická
 - Aritm. Priemer = medián





Age of Nobel Prize Winners



Normálne rozdelenie

Histogram spojitej náhodnej premennej, ktorá má normálne rozdelenie, má vždy základný tvar gaussovej krivky, ale mení sa v závislosti od parametrov: aritmetického priemeru μ a smerodajnej odchýlky σ .

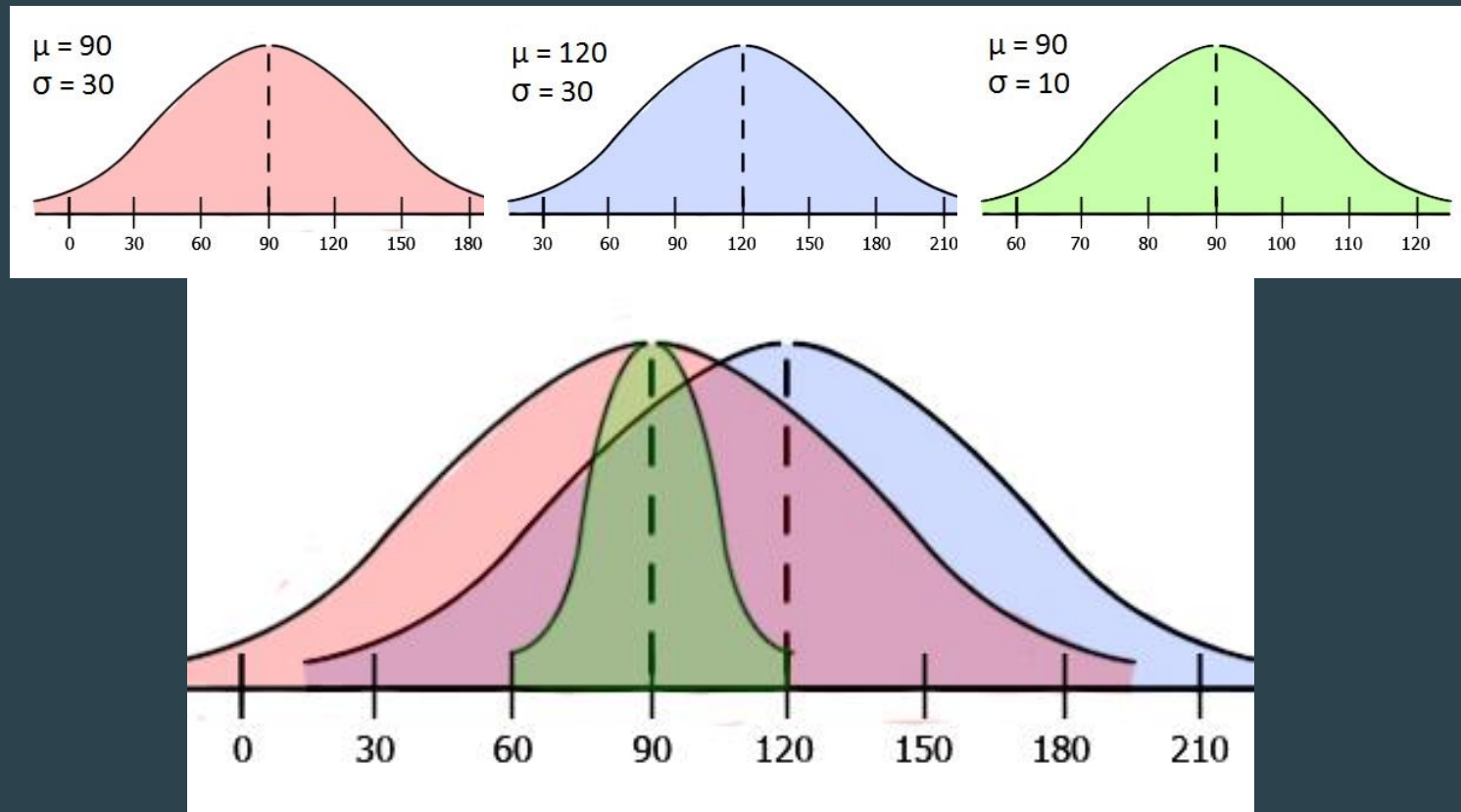
Zostrojme histogram premennej s nasledujúcimi parametrami:

a) $\mu = 90, \sigma = 30$

b) $\mu = 120, \sigma = 30$

c) $\mu = 90, \sigma = 10$

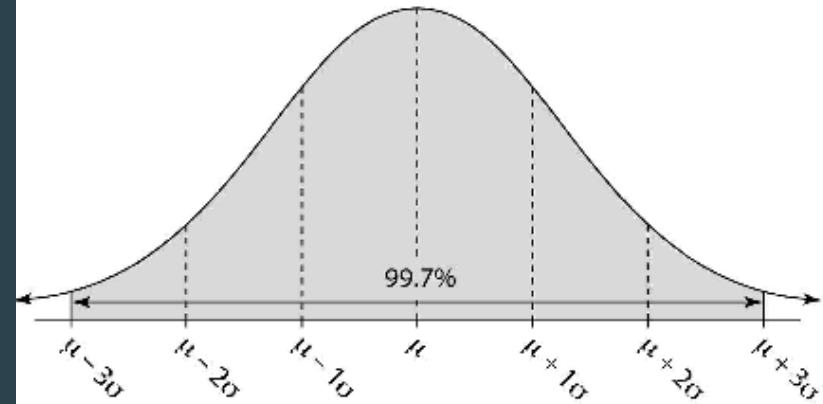
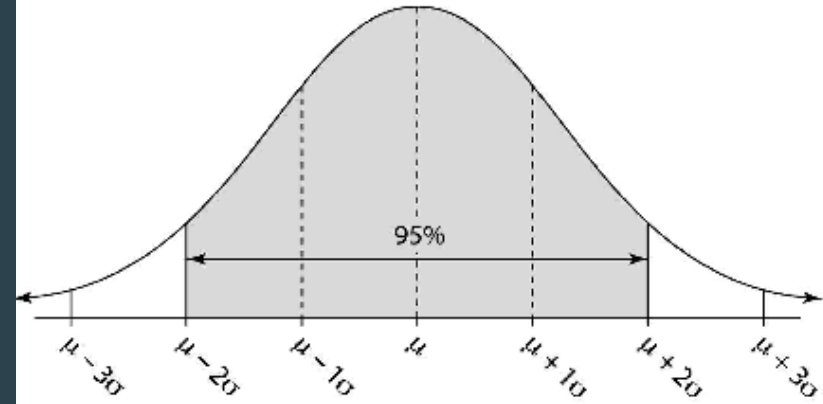
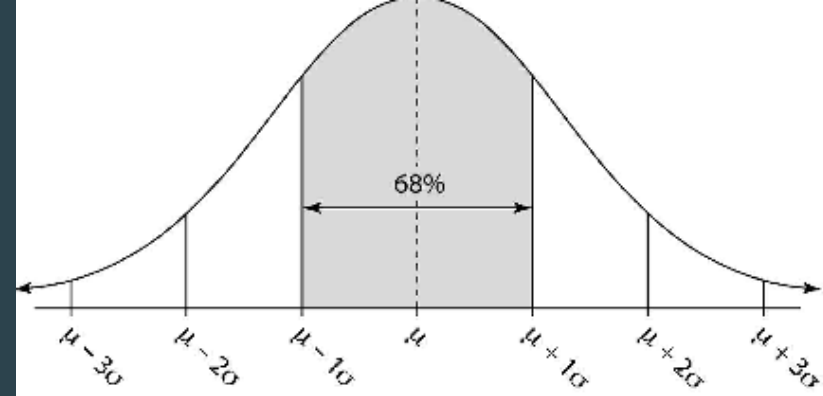
Grafy normálneho rozdelenia



Normálne rozdelenie

- Hodnota aritmetického priemeru μ posúva gaussovu krivku vpravo alebo vľavo po číselnej osi
- Hodnota smerodajnej odchýlky σ mení tvar krivky: čím je odchýlka väčšia, tým je krivka „tučnejšia“

Pravidlo troch sigma (3σ) alebo pravidlo 68-95-99,7 (približné hodnoty)



Príklad

- V teste [SAT](#) Anna získala 1800 bodov a Tom v teste [ACT](#) získal 24 bodov.

Z dlhodobých záznamov vieme, že počet bodov v teste SAT má normálne rozdelenie s $\mu_1 = 1500$ a $\sigma_1 = 300$, počet bodov ACT má normálne rozdelenie s $\mu_2 = 21$ a $\sigma_2 = 5$.

Kto dopadol v teste lepšie? Anna alebo Tom?

Z-skóre

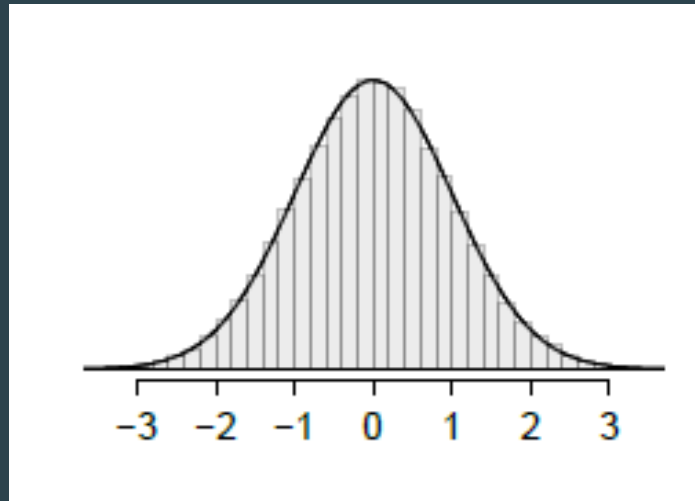
- Pre každú hodnotu spojitej náhodnej premennej môžeme vypočítať z-skóre, ktoré predstavuje o aký násobok smerodajnej odchýlky sa daná hodnota odchýlila od aritmetického priemeru.
- Ak x je hodnota premennej, ktorá má normálne rozdelenie s priemerom μ a odchýlkou σ , tak z-skóre hodnoty x je definované:

$$Z = \frac{x - \mu}{\sigma}$$

- `=standardize(x;μ;σ)`

Štandardné normálne rozdelenie

- = normálne rozdelenie s parametrami: aritmetický priemer $\mu = 0, \sigma = 1$



Percentil

- opisná štatistika vyjadrujúca relatívne postavenie konkrétnej hodnoty;
- k -ty percentil rozdeľuje všetky dáta (usporiadané vzostupne) na dve časti: jedna časť obsahuje $k\%$ všetkých dát, druhá časť zvyšných $(100 - k)\%$ dát
- medián = 50-ty percentil
- prvý kvartil = Q_1 = 25-ty percentil
- tretí kvartil = Q_3 = 75-ty percentil
- k je číslo medzi 1 a 100
- typickým príkladom je výsledok testu na prijímacích skúškach, SCIO alebo monitory

Table 5-4 Percentiles for All Possible ACT Exam Scores in Math and Reading		
<i>ACT Score</i>	<i>Mathematics Percentile</i>	<i>Reading Percentile</i>
34–36	99	99
33	98	97
32	97	95
31	96	93
30	95	91
29	93	88
28	91	85
27	88	81
26	84	78
25	79	74
24	74	70
23	68	65
22	62	59
21	57	54
20	52	47
19	47	41
18	40	34
17	33	30
16	24	24
15	14	19
14	06	14
13	02	09
12	01	06
11	01	03
1–10	01	01

Príklad

- Tomáš získal v testoch ACT z matematiky a aj z čítania s porozumením zhodou okolností 30 bodov. V ktorom teste dopadol relatívne lepšie?

Príklad

Anna z predchádzajúceho príkladu by chcela vedieť, aký percentil dosiahla pri svojom zisku 1800 bodov z testu SAT, t.j. koľko percent študentov dopadlo horšie, resp. lepšie ako ona.

Tabuľka normálneho rozdelenia

Nám umožňuje ku každej hodnote Z-skóre nájsť jej príslušný percentil a naopak.

Napr. ak $Z=0,43$, tak v tabuľke nájdeme riadok 0,4 a stĺpec 0,03 a ich priesečníku nájdeme percentil normálneho rozdelenia ku Z-skóre 0,43

`=norm.dist(x;μ;σ>true)`

`=norm.inv(percentil;μ;σ)`

Tabuľka normálneho rozdelenia

[illegible]

Príklady

Počet bodov v teste SAT má normálne rozdelenie s aritmetickým priemerom 1500 a smerodajnou odchýlkou 300.

1. Aká je pravdepodobnosť, že náhodne vybraný študent získa aspoň 1630 bodov?
2. Larry verí, že na svoju vysnívanú univerzitu sa dostane, ak v teste SAT dosiahne 80-ty percentil. Koľko bodov musí získať?

Príklad 1 – riešenie

Aká je pravdepodobnosť, že náhodne vybraný študent získa aspoň 1630 bodov?

Príklad 1 – riešenie

Príklad 2 – riešenie

Larry verí, že na svoju vysnívanú univerzitu sa dostane, ak v teste SAT dosiahne 80-ty percentil. Koľko bodov musí získať?

Tabuľka normálneho rozdelenia

[illegible]

Príklad 2 – riešenie

Príklad 2 – riešenie