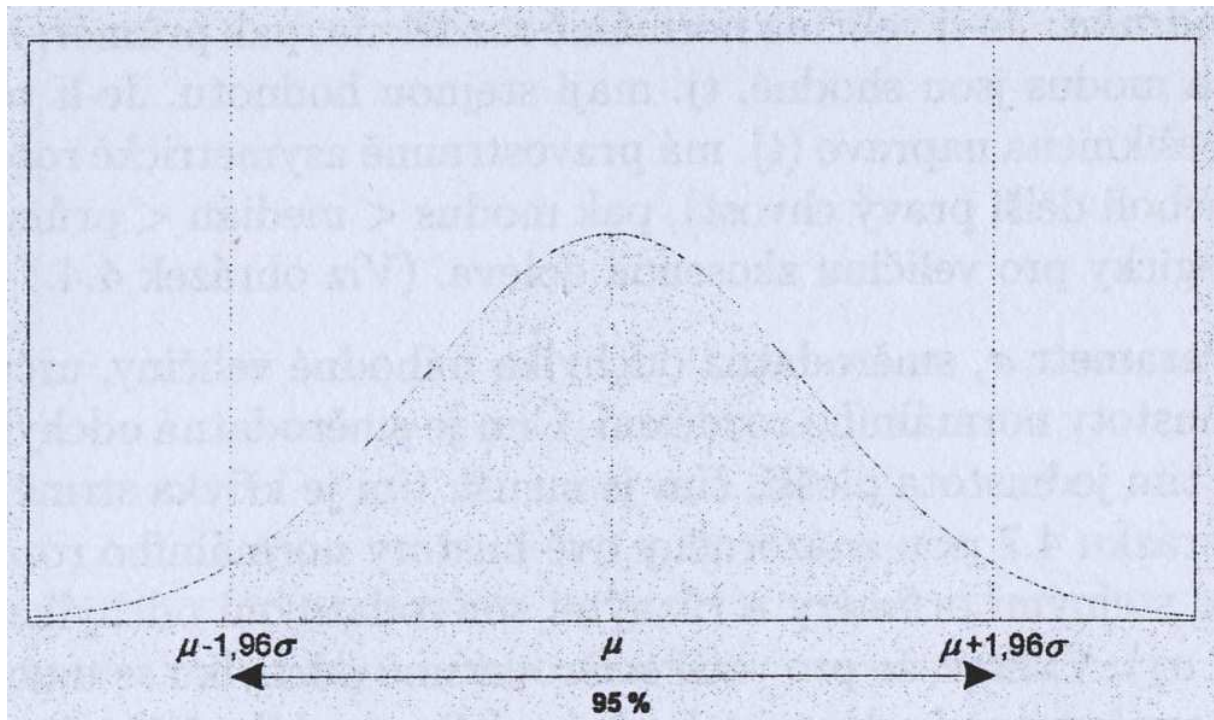


Normální rozdělení dat



Pokud máme dostatečný počet výsledků testování, či jiných dat, v obecném případě bude jejich grafické rozložení podobné znázornění na obrázku.

Tzn. že nejčastěji se budou vyskytovat hodnoty blízké průměrné hodnotě. Čím více se hodnoty liší od průměru, tím méně často zastoupené jsou.

Toto rozdělení se často podle svého objevitele nazývá Gaussovo rozdělení.

MÍRY POLOHY

Aritmetický průměr \bar{x} , μ

- součet všech hodnot dělíme rozsahem souboru

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- tento znak je velmi citlivý na odlehlé hodnoty

Medián \tilde{x} , MED

- prostřední hodnota výběrového souboru
- u sudého počtu měření se jedná o průměr ze dvou prostředních hodnot

Modus \hat{x} , MOD

- nejčastěji se vyskytující veličina
- užívány zejména u kvalitativních nominálních znaků

MÍRY VARIABILITY

Ukazují, jak daleko jsou hodnoty vzdáleny od centra.

Variační rozpětí R

- rozdíl mezi nejvyšší a nejnižší hodnotou dat

$$R = x_{\max} - x_{\min}$$

- značně závislé na extrémních hodnotách

Rozptyl s^2, σ^2

- průměr čtverců odchylek od průměru

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

Směrodatná odchylka s, σ

- v praxi nejčastěji používaný ukazatel variability
- její výhoda je, že má stejné jednotky jako měřená veličina
- jedná se o odmocninu rozptylu

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \cong \frac{1}{6} R$$

Normované testové výsledky

Výsledky získané v jednotlivých testech jsou vyjádřeny v různých jednotkách (fyzikální – m, s, ..., body, počet opakování, apod.)

Abychom je mohli porovnávat, sčítat, vyhodnocovat, převádíme je na tzv. **normované body**. Nejznámější z normovaných stupnic jsou:

z-body
T-body
percentily
steny

K vytvoření normované stupnice je třeba otestovat poměrně velké soubory ($n > 100$), z jejichž výsledků normu vytvoříme.

z-body

odchylku testového výsledku od průměru normové populace dělíme směrodatnou odchylkou

$$z = \frac{x_i - \bar{x}}{s}$$

x_i ... konkrétní výkon
 \bar{x} ... aritmetický průměr
 s ... směrodatná odchylka

Příklad:

Skok z místa – výkon 213, průměrný výkon 205, směrodatná odchylka 11,5

$$z = \frac{x_i - \bar{x}}{s} = \frac{213 - 205}{11,5} \cong 0,7$$

Protože výsledky na z-bodové stupnici jsou pro praktické použití nevhodné (malá čísla, záporné hodnoty), používáme je jako základ pro výpočet T-bodů.

T-body

$$T = 50 + 10z$$

50 ... konstanta
z ... výsledek z-bodu

Příklad:

$$T = 50 + 10z = 50 + 10 * 0,7 = 57$$

Konstrukce T-bodové stupnice:

- Vypočteme hodnoty \bar{x} a s referenčního souboru
- Přidělíme hodnotě \bar{x} 50 bodů
- postupujeme ve stupnici směrem nahoru (51, 52, ...) a dolů (49, 47, ...) tak, že **$1b = s / 10$**
- tuto „teoretickou“ stupnici upravíme na prakticky měřitelné hodnoty

Výpočet a využití percentilů

Nejjednodušší způsob zhodnocení výkonu testované osoby v některém testu je určení jejího pořadí v souboru (pracujeme s tzv. pořadovou stupnicí)

Např. žák dosáhne 3. nejlepšího výkonu mezi 35 spolužáky ve třídě, což jej srozumitelně vyjádření jeho pozice ve skupině. Ale být 3. mezi 35 nebo 3. mezi 100 žáky je podstatný rozdíl.

Proto je výhodnější vyjádřit pořadí v procentech. Provádíme tedy měření tzv. **procentilového skóre**, zkráceně **procentilu (percentilu)**.

Percentil určuje relativní pozici testované osoby ve skupině, informuje nás o tom, kolik procent osob skóruje níže než testovaná osoba.

Naměřený výkon převedeme na percentily podle vzorce:

$$P = \frac{kumf - 0,5}{n} * 100$$

kumf...kumulativní četnost – narůstá s pořadím výkonů
ve skupině od nejhoršího po nejlepší
n...počet testovaných osob

Příklad:

$$P = \frac{kumf - 0,5}{n} * 100 = \frac{33 - 0,5}{35} * 100 \cong 93$$

Testovaná osoba dosahuje 93. percentilu tzn., že 93% žáků je horších, 7% lepších.

Padesátý percentil – **medián**

Skóre, které dosahuje prostřední člen skupiny.

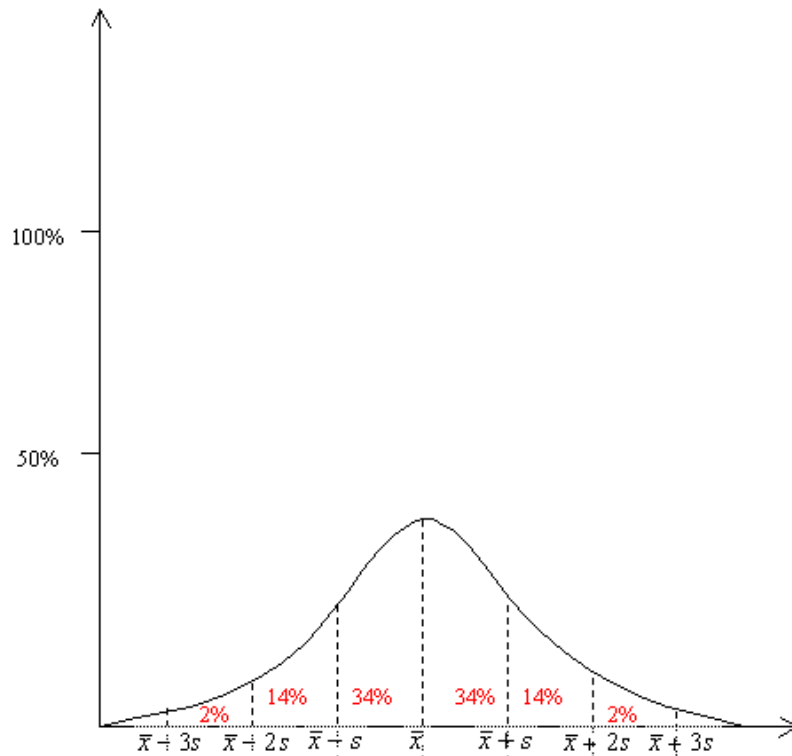
Za předpokladu normálního rozložení četností se medián rovná aritmetickému průměru.

Percentilové stupnice (pro grafické znázornění výsledku)

Konstruují se z výsledků velkých souborů (stovky – tisíce testovaných osob)

Nemá charakter pravidelné ekviintervalové stupnice.

Vychází z kumulativních četností jednotlivých hrubých skóre → kolem mediánu je stupnice „hustší“, ke krajům „řídne“.



0.1 2 16 50 84 98 99.9 P = procentily

-3 -2 -1 0 1 2 3 Z = body

20 30 40 50 60 70 80 T-body = 50+10Z

2%	5%	9%	15%	19%	19%	15%	9%	5%	2%
1	2	3	4	5	6	7	8	9	10

ST-body(STENY) = 50+10Z

55 70 85 100 115 130 145 MQ.body = 100 + 15Z

Testování statistických hypotéz

Jejich úkolem je ověřovat oprávněnost určitých předpokladů (hypotéz) vyslovených o vlastnostech dějů.

Podstata testování hypotéz je předpoklad platnosti určité okolnosti (tzv. nulové hypotézy H_0), respektive její neplatnosti (\rightarrow přijetí tzv. alternativní hypotézy H_a).

Riziko chyby přijetí – zamítnutí H_0 se nazývá **hladina významnosti**, značí se písmenem α . Zpravidla volíme $\alpha = 0,05$ nebo $0,01$ (tj. 5% nebo 1% riziko pravděpodobnosti).

Snížení rizika chybného rozhodnutí nejsnáze dosáhneme zvětšováním rozsahu výběru souboru.

Testovací kritérium se označuje t .

Jestliže vypočtené t je menší než tzv. kritická hodnota, přijímáme H_0 (rozdíl není statisticky významný).

Jestliže vypočtené t překračuje tabulkovou hodnotu, platí H_a (rozdíl je statisticky významný – signifikantní).

Pro výpočet t-testu používáme různé vzorce pro tyto případy:

a) tzv. **závislé soubory** (párové hodnoty)
jde o testování stejných probandů s určitým časovým odstupem

b) tzv. **nezávislé soubory**
jedná se o soubory různých osob

Postup:

- vypočteme \bar{x} , s pro 1. a 2. soubor
- tzv. F-testem zjistíme, zda jde o „shodné“ nebo „rozdílné“ rozptyly
- použijeme příslušný vzorec pro shodné nebo rozdílné soubory