

PSY117/454

Statistická analýza dat v psychologii

**Přednáška 2**

---

# ČETNOSTI A ROZLOŽENÍ ČETNOSTÍ

Je snadné lhát s pomocí statistiky. Je těžké říkat pravdu bez ní.

*Andrejs Dunkels; wikiquote*

# Tabulka četností (frekvencí)

<b>hodnota/ interval</b>	<b>(absolutní) četnost</b>	<b>kumulativní četnost</b>	<b>relativní četn. (%)</b>	<b>kumulativní rel. č.</b>
hodnota1/ interval1				
hodnota2/ interval2				
...				
maximum/ posl. interv.		N		100
<b>Celkem</b>	<b>N</b>		<b>100</b>	

Fuj: „počet“ v Tab 3.2, hustota (jde o hustotu pravděpodobnosti), obr. 3.5 – ne frekvence, ale procenta  
 AJ: (absolute) frequencies, relative frequencies, percent, cumulative, value, interval, total, N=sample size

# Tabulka četností - poznámky

---

- od nejmenší hodnoty po nejvyšší
- v 1. a 2. sl. obvykle zahrnuty chybějící hodnoty
  - Pak se rozlišuje mezi platnými hodnotami a chybějícími hodnotami
- hodnoty – diskrétní proměnné
- intervaly – spojité proměnné
  - volba velikosti intervalu
    - aby byl jejich počet přibližně  $N/10$  nebo  $<15$
    - stejná šířka všech intervalů
- pojem odlehlá hodnota (outlier)

# Grafické podoby tabulky četností

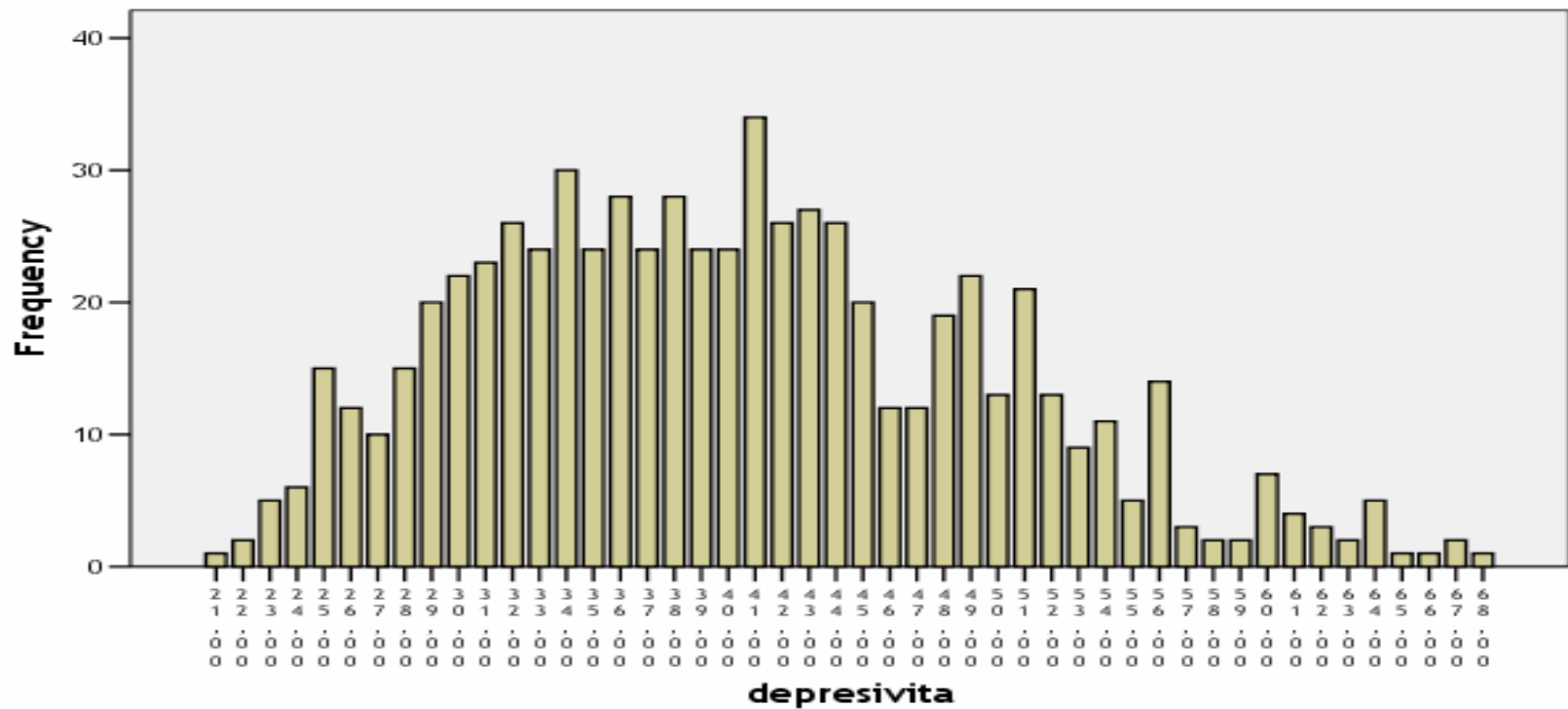
---

- ❑ sloupcový graf (diagram) – nominální, ordinální
- ❑ histogram – intervalové, poměrové
- ❑ koláčový diagram – zřídka, neukazuje rozložení
- ❑ Každý graf (i tabulka) musí být natolik přehledně popsán (nadpis + popisky uvnitř), aby byl srozumitelný i bez čtení textu
- ❑ *Když nelžou čísla, lžou grafy ...*

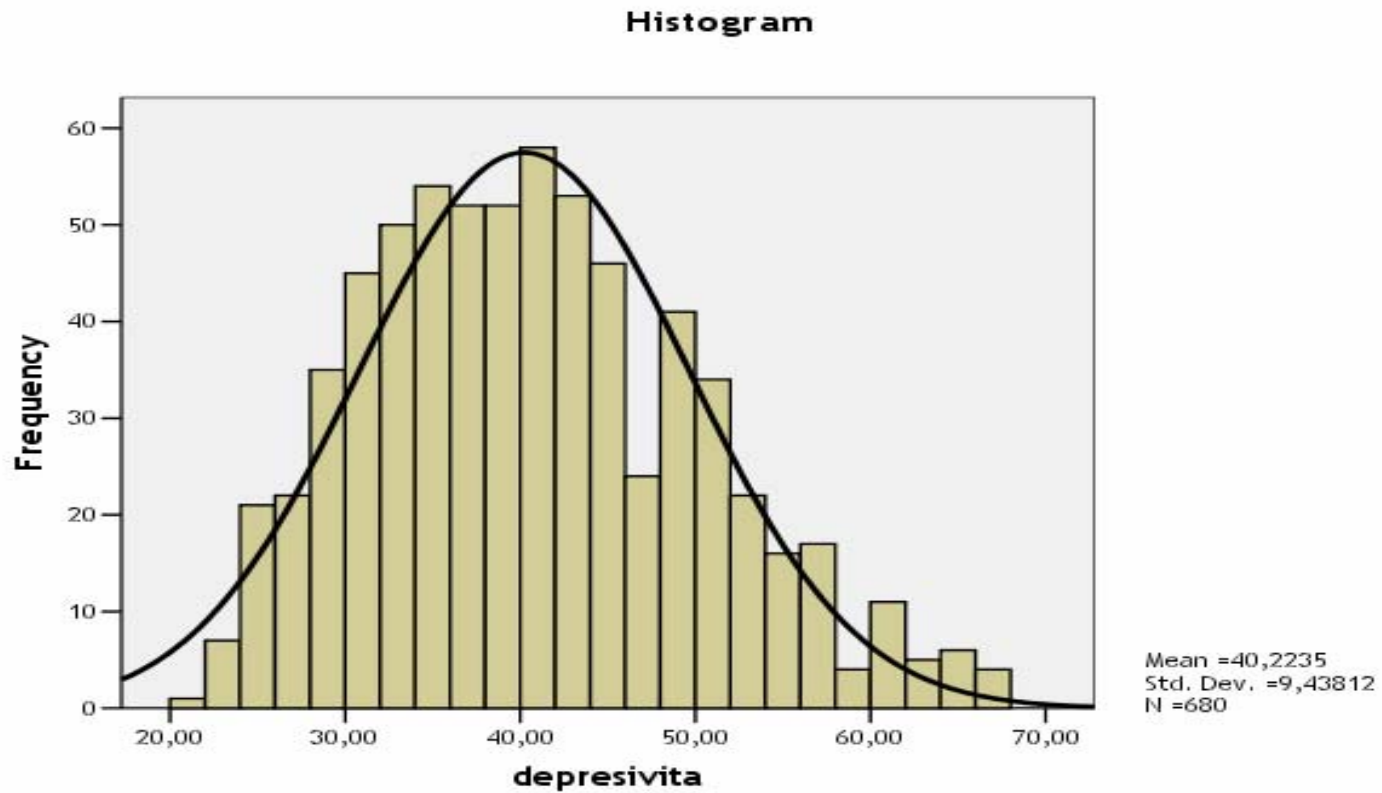
[http://en.wikipedia.org/wiki/Misuse\\_of\\_statistics](http://en.wikipedia.org/wiki/Misuse_of_statistics)

# Sloupcový diagram

depresivita

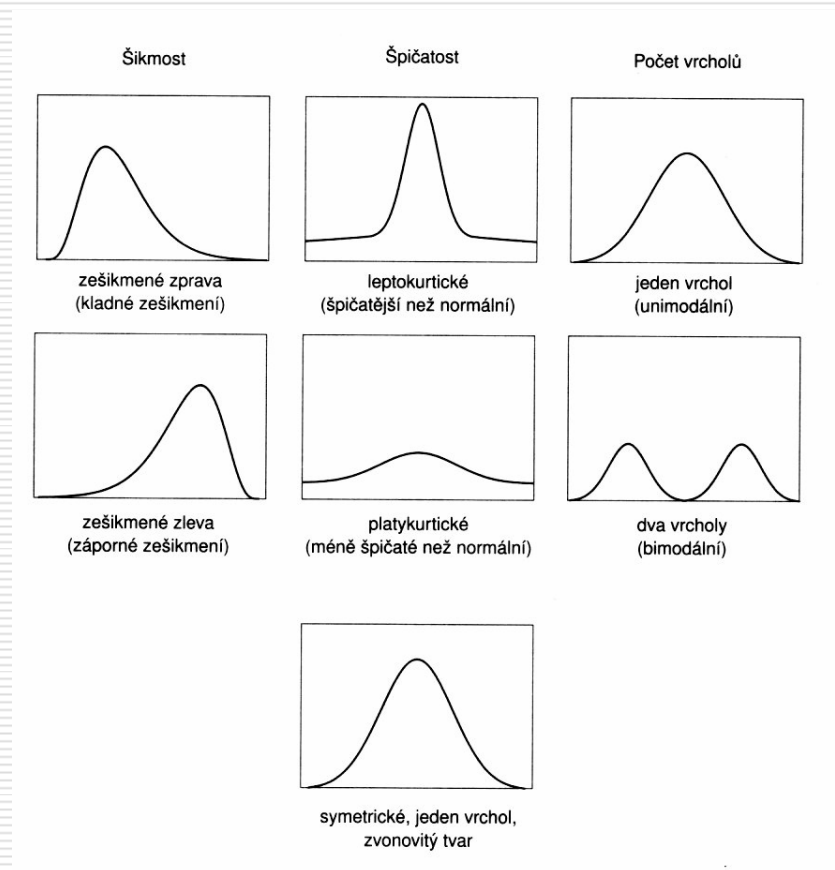


# Histogram

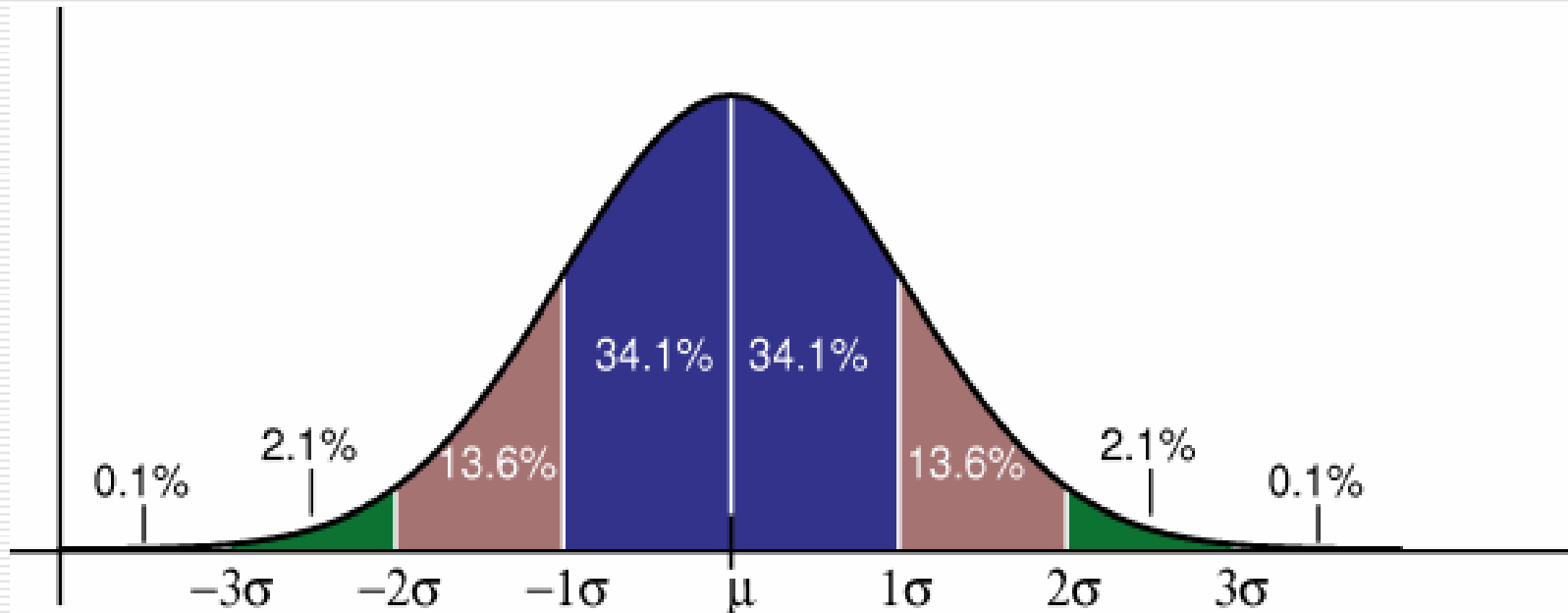


# Rozložení *rozdělení, distribuce* četností

- Normální
- Uniformní
- Počet vrcholů
  - Unimodální, bimodální, multimodální
- Zešikmení
  - Zešikmené zprava (pozitivně), efekt podlahy
  - Zešikmené zleva (negativně), efekt stropu
- Strmost
  - Leptokurtické, platykurtické



# Normální (Gaussovo) rozložení

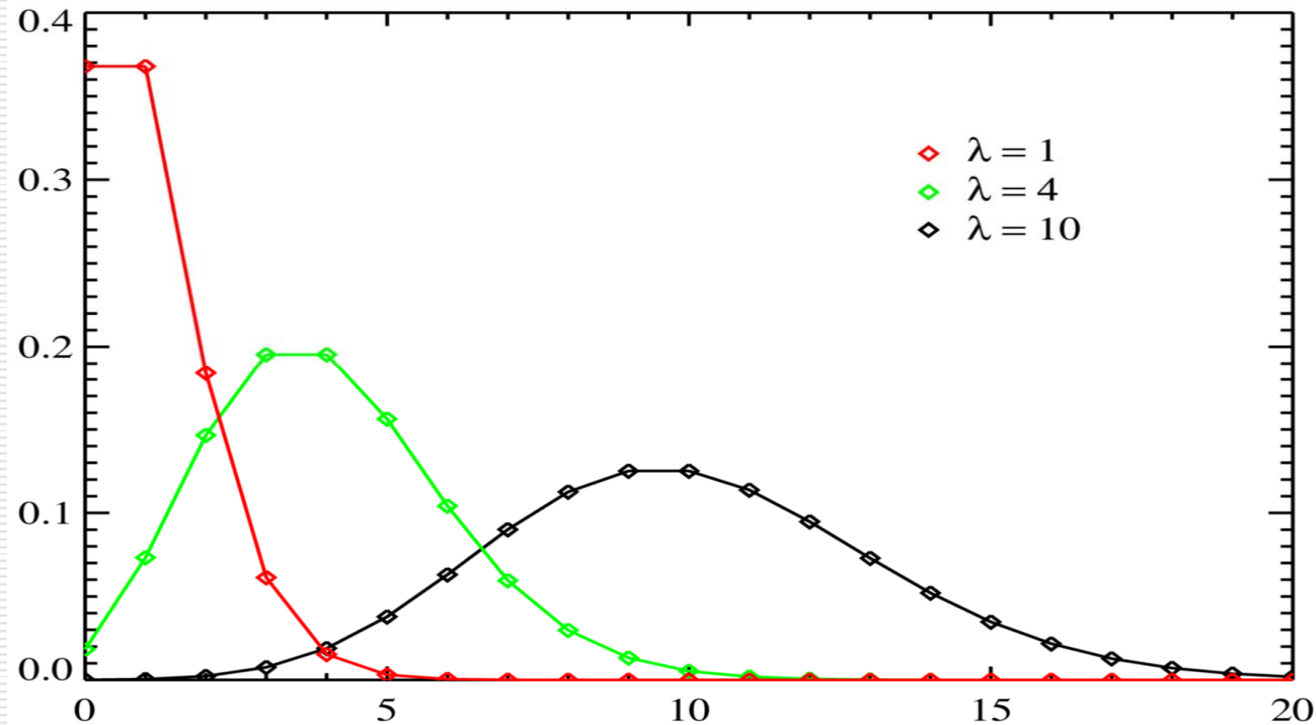


[http://en.wikipedia.org/wiki/Image:Standard\\_deviation\\_diagram.png](http://en.wikipedia.org/wiki/Image:Standard_deviation_diagram.png)

- ❑ „Normální“ ve smyslu „velmi běžné“
- ❑ Tam, kde se setkává mnoho nezávislých vlivů.
- ❑ Ne vždy, nesouvisí s „kvalitou“ dat.



# Poissonovo rozložení



- Rozložení řídkých událostí (ta lambda v grafu = frekvence za jednotku času)
- Děje-li se událost častěji, než 10x za časovou jednotku, která nás zajímá, je jeho dobrou aproximací normální rozložení.

# Shrnutí

---

- ❑ První informací *statistikou*, která nás zajímá je **četnost** výskytu jednotlivých hodnot (resp. hodnot uvnitř jednotlivých intervalů)
- ❑ Konfiguraci **četností** nazýváme **rozložení**.
- ❑ O rozložení hodnot proměnné uvažujeme většinou graficky – **histogram, sloupcový diagram**.
- ❑ Nej... rozložením je tzv. **normální rozložení**.
- ❑ Byť tohle je 5. třída ZŠ – už tady se **podvádí**.