

# Regresní analýza – souhrn (co je potřeba si pamatovat)

- Pomocí regresní analýzy (dále RA) zkoumáme závislost 2 a více číselných proměnných a odhadujeme hodnoty závisle proměnné  $y$  odpovídající dané hodnotě jedné či více nezávisle proměnných.
- RA se zabývá **jednostrannými** závislostmi (tj. závislost  $y$  na  $x$ , ne pak už naopak).
- Jestliže pracujeme s 1 závisle proměnnou (vysvětlovanou proměnnou)  $y$  a jednou nezávisle proměnnou (vysvětlující proměnnou)  $x$ , hovoříme o jednoduché (párové) regresní analýze.
- Jestliže pracujeme s 1 závisle proměnnou (vysvětlovanou proměnnou)  $y$  a dvěma či více nezávisle proměnnými (vysvětlujícími)  $x_1, x_2, \dots, x_k$ , hovoříme o vícenásobné regresní analýze.
- Postup regresní analýzy:
  1. Volba typu regresní funkce.
  2. Odhad parametrů zvolené regresní funkce.
  3. Testování hypotéz o parametrech regresní funkce.
  4. Ověření vhodnosti vybraného regresního modelu.

## 1. Volba typu regresní funkce

- Snažíme se vybrat funkci co nejjednodušší, se snadno interpretovatelnými parametry.
- Vybraná regresní funkce by měla co nejtěsněji přiléhat napozorovaným hodnotám, což znamená, že se snažíme, aby odchylky skutečných hodnot od hodnot vyrovnaných byly co nejmenší.
- Způsob výběru vhodné regresní funkce: a) bodový diagram, b) matematicko-statistická kritéria (viz studijní materiál k tématu 4, Kritéria pro posouzení kvality zvolené regresní funkce), c) věcně-ekonomická kritéria.

## 2. Odhady parametrů zvolené regresní funkce

- Základní úvaha: máme definovanou **teoretickou** (hypotetickou) regresní funkci, která je nezměřitelná, a na druhé straně **empirickou** (výběrovou) regresní funkci, která je zkonstruována na základě empirických údajů. Empirickou regresní funkci považujeme za odhad teoretické regresní funkce.
- Metody odhadu parametrů teoretické regresní funkce volíme podle toho, zda je funkce **lineární** nebo **nelineární** ve svých parametrech.
- Mezi funkce lineární v parametrech patří: regresní přímka, regresní parabola, regresní hyperbola, regresní rovina, regresní logaritmická funkce a další.
- Mezi funkce nelineární v parametrech patří: regresní mocninná funkce, regresní exponenciální funkce a další.
- Odhad parametrů funkcí lineárních v parametrech provádíme pomocí **metody nejmenších čtverců**, která spočívá v minimalizaci součtu čtverců náhodných chyb, tj. minimalizujeme součet čtverců odchylek empirických hodnot od hodnot vyrovnaných.  
*Poznámka:* V praxi odhady parametrů jakýchkoli funkcí ručně už neprovádíme, jsou již přednastaveny v používaných statistických programech.

- Odhad parametrů funkcí nelineárních v parametrech nelze provést pomocí metody nejmenších čtverců, anebo ne přímo. Obvykle najdeme nějaký vhodný počáteční odhad a postupným zlepšováním řešení pomocí iteračních postupů je nalezen odhad s požadovanou přesností. Mezi metody počátečního odhadu patří např. metoda linearizující transformace, metoda vybraných bodů, metoda apriorní informace (úkol: najděte v literatuře podrobnosti o těchto metodách).

### 3. Testování hypotéz o parametrech regresní funkce

- Testy nazýváme individuální t-testy o nulové hodnotě regresních parametrů.
- Testují, zda parametry teoretické regresní funkce nejsou statisticky významné, tj. postupně testujeme, že parametry  $\beta_0, \beta_1, \dots, \beta_k$  jsou rovny 0.
- Snažíme se, aby byla prokázána alternativní hypotéza, která říká, že daný parametr je statisticky významný, tj. že má pro danou funkci přínos.

### 4. Ověření vhodnosti zvoleného regresního modelu

- Provádí se pomocí celkového F-testu (je založen na rozkladu rozptylu vysvětlované proměnné).
- $H_0$  zde předpokládá, že zvolená regresní funkce **není** vhodná k popisu dané závislosti.  $H_1$  naopak tvrdí, že se vybraná funkce k popisu zkoumané závislosti hodí.
- **Závěry o vhodnosti** zvoleného regresního modelu vycházejí primárně z výsledku celkového F-testu. Pokud celkový F-test vede k zamítnutí  $H_0$ , říkáme o něm, že je významný (totéž říkáme o t-testech, pokud vedou k zamítnutí  $H_0$ ).
  - a) Pokud je celkový F-test nevýznamný (tj. nevede k zamítnutí  $H_0$ ), není vhodné vybranou regresní funkci k popisu zkoumané závislosti používat, a to ani tehdy, jsou-li některé z t-testů významné.
  - b) Pokud je celkový F-test významný a většina t-testů rovněž, považujeme danou regresní funkci za vhodnou k popisu zkoumané závislosti.
  - c) Pokud je celkový F-test významný a většina t-testů ne, je potřeba se zamyslet, zda daná regresní funkce nemá příliš velký počet parametrů, a zda by nebylo vhodné některé z nich vypustit, tj. zvolit jednodušší model.
- Pokud se stane, že průběh některé závislosti je možné popsat pomocí více regresních funkcí (tj. více funkcí má významný F-test a většinu t-testů), je potřeba vybrat nejvhodnější z těchto funkcí. Mezi kritéria pro posouzení kvality regresní funkce patří: index determinace, testové kritérium F z celkového F-testu, reziduální součet čtverců, reziduální rozptyl.