Metody prognózování v dopravě

Ing. Michal Dorda, Ph.D.

Metody prognózování v dopravě

- Cílem prognózy dopravy je určení výhledových údajů o dopravě (např. výhledové intenzity dopravy apod.).
- Při prognózování v dopravě je užívána celá řada metod – analýza časových řad, regresní a korelační analýza, metody koeficientů růstu, gravitační metody atd.

Analýza trendu časové řady

- Nejjednodušším způsobem prognózy je **extrapolace dosavadních dat**.
- Mějme sledované údaje seřazené v **časové řadě**. Na základě analýzy této časové řady (**analýza trendu časové řady** apod.) jsme schopni extrapolovat hledané údaje pro výhledové období.

Analýza trendu časové řady

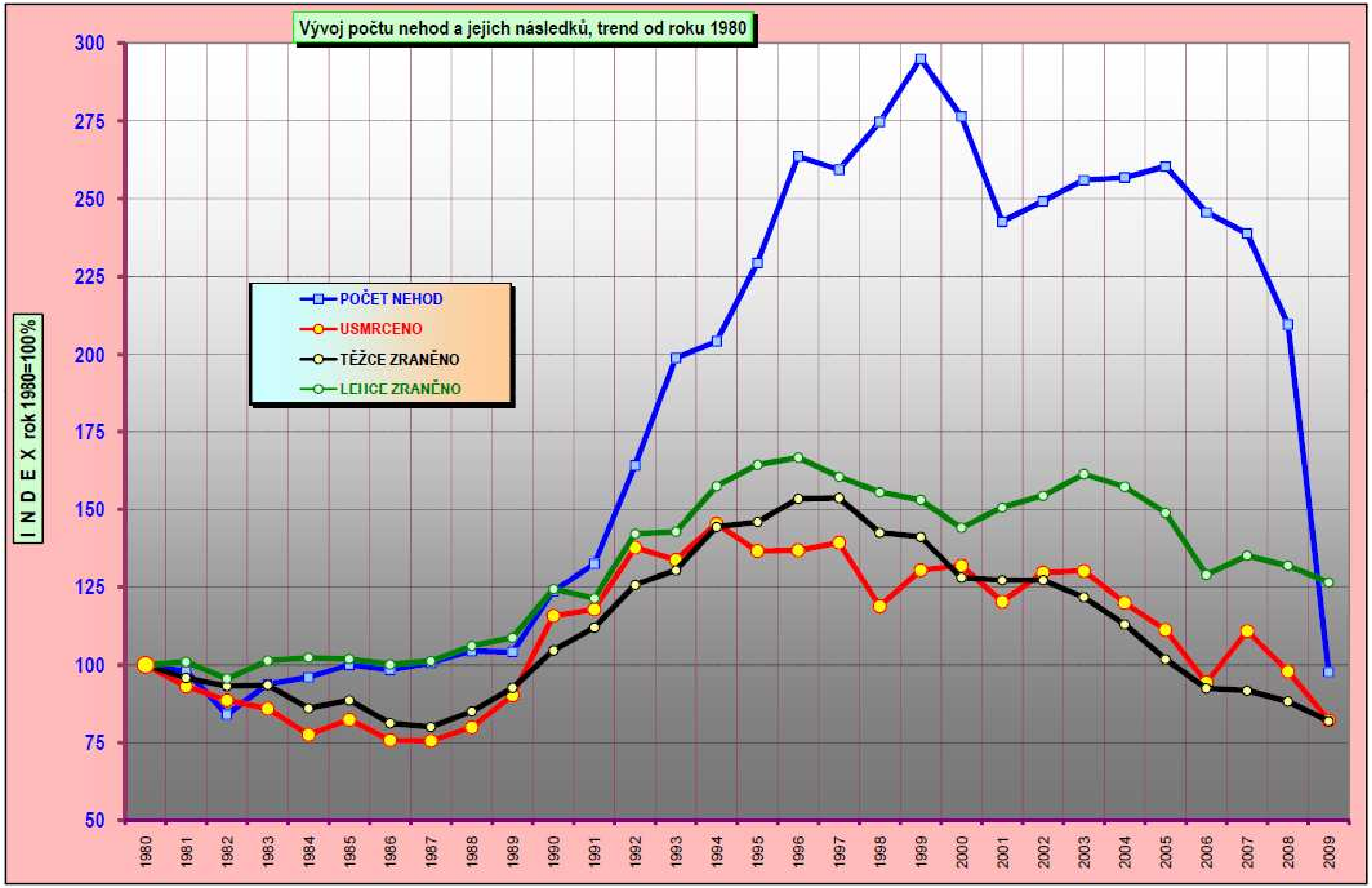
- Např. známe-li vývoj intenzit na pozemní komunikaci z předchozích období, jsme schopni na základě analýzy trendu této časové řady odhadnout výhledové intenzity.
- Z tohoto důvodu se musíme seznámit se základními principy analýzy trendu časových řad.

Úvod do časových řad, analýza trendu časové řady

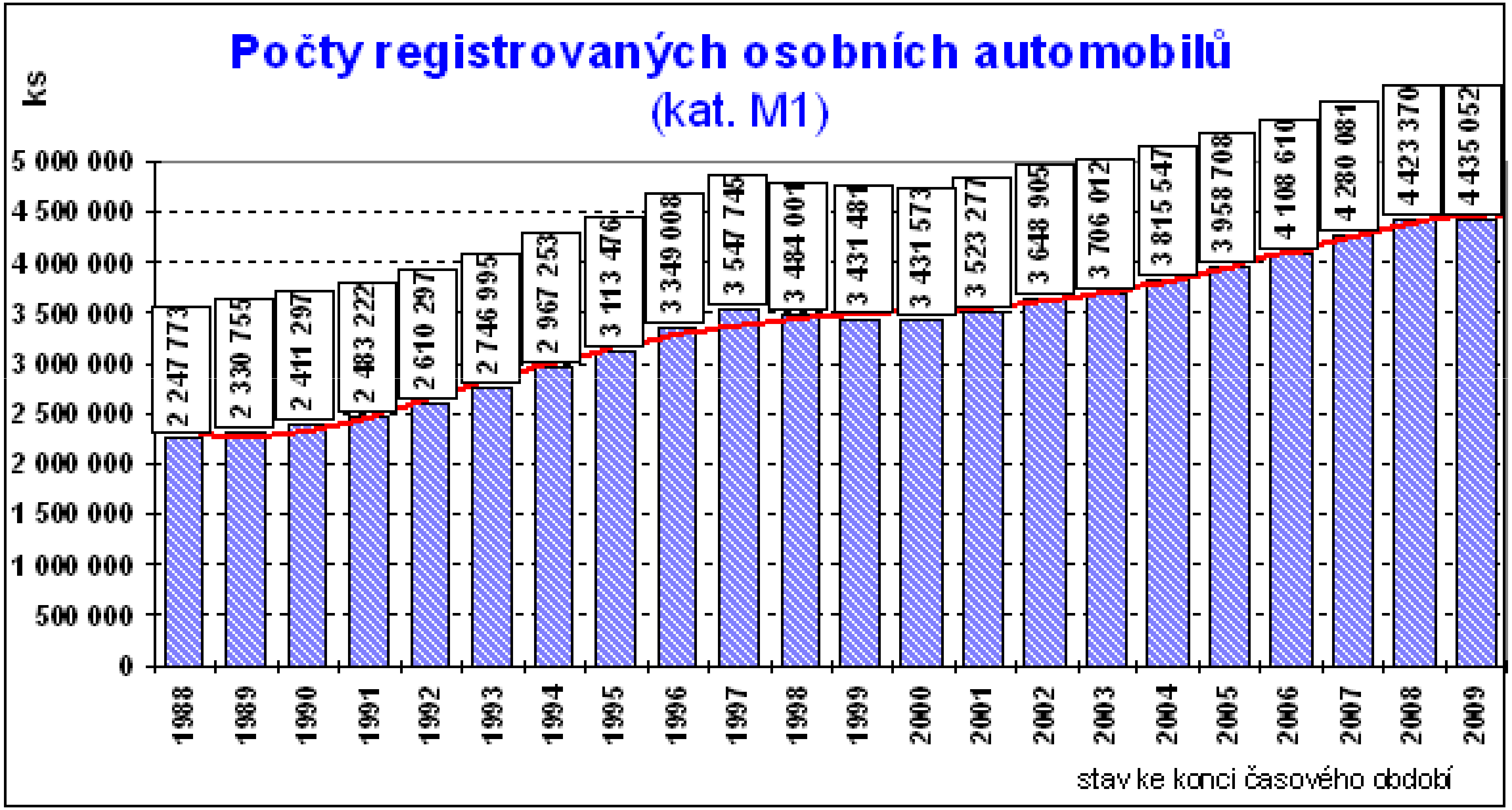
Úvod do časových řad

- **Časovou řadou** rozumíme posloupnost věcně a prostorově srovnatelných pozorování (dat), která jsou jednoznačně uspořádána z hlediska času ve směru minulost – přítomnost.
- Časovou řadou v oblasti dopravy může být např. počet dopravních nehod v jednotlivých letech, počet registrovaných vozidel v jednotlivých letech, vývoj intenzit na pozemní komunikaci apod.

Úvod do časových řad



Úvod do časových řad



Úvod do časových řad

- Časové řady lze členit podle různých kritérií:
    1. Podle rozhodného časového hlediska.
    2. Podle periodicity, s jakou jsou údaje sledovány.
    3. Podle druhu sledovaných ukazatelů.
    4. Podle způsobu vyjádření údajů.

Úvod do časových řad

1) Podle rozhodného časového hlediska rozlišujeme:
   - **Časové řady intervalové** (resp. časové řady intervalových ukazatelů).
   - **Časové řady okamžikové** (resp. časové řady okamžikových ukazatelů).

Úvod do časových řad

- **Intervalovou časovou řadou** rozumíme řadu takového ukazatele, jehož velikost závisí na délce intervalu, za který je sledován.
  – Pro ukazatele tohoto typu má smysl tvořit součty, ukázkou intervalové časové řady může být řada zobrazující vývoj počtu dopravních nehod v jednotlivých letech.

Úvod do časových řad

- Sledované údaje se mají vztahovat ke stejně dlouhým časovým intervalům, provádíme tzv. **očištění časových řad od vlivů kalendářních variací** (sledované údaje přepočítáváme na jednotkový časový interval)

Úvod do časových řad

- Údaje očištěné na **kalendářní dny** získáme podle vztahu:

$$y_t^{(o)} = y_t \cdot \frac{\bar{k}_t}{k_t},$$

kde: y_t je hodnota očišťovaného ukazatele,

k_t je počet kalendářních dní v daném období,

$\bar{k}_t$ je průměrný počet kalendářních dní v daném období.

Úvod do časových řad

$$\bar{k}_t = \frac{\sum_{t=1}^{n} k_t}{n} = \frac{366}{12} = 30{,}5$$

Měsíc	Počet dní měsíce k_t	Počet nehod y_t	Očištěný počet nehod $y_t^{(0)}$
Leden	31	18 939	18 634
Únor	29	16 137	16 972
Březen	31	17 849	17 561
Duben	30	15 724	15 986
Květen	31	17 694	17 409
Červen	30	17 914	18 213
Červenec	31	16 699	16 430
Srpen	31	17 386	17 106
Září	30	16 829	17 109
Říjen	31	19 105	18 797
Listopad	30	18 644	18 955
Prosinec	31	18 596	18 296

např. $y_1^{(0)} = y_1 \cdot \frac{\bar{k}_t}{k_1} = 18939 \cdot \frac{30{,}5}{31} \doteq 18634$

Úvod do časových řad

- Údaje očištěné na **pracovní dny** získáme podle vztahu:

$$y_t^{(o)} = y_t \cdot \frac{\overline{p}_t}{p_t},$$

kde: y_t je hodnota očišťovaného ukazatele,

p_t je počet pracovních dní v daném období,

$\overline{p}_t$ je průměrný počet pracovních dní v daném období.

Úvod do časových řad

- **Okamžikové časové řady** jsou tvořeny z údajů, které se vztahují k určitému okamžiku.
  - Příkladem může být počet evidovaných vozidel v ČR k 31. 12. každého roku.
  - U těchto řad nemá smysl stanovovat součty.

Úvod do časových řad

2) Podle periodicity, s jakou jsou údaje sledovány, rozlišujeme:
   - **Krátkodobé časové řady** (periodicita je kratší než 1 rok) – zpravidla 1 měsíc.
   - **Roční (dlouhodobé) časové řady** (periodicita je roční nebo ještě delší).

Úvod do časových řad

3) Podle druhu sledovaných ukazatelů rozlišujeme:
   - **Časovou řadu absolutních hodnot** (zpravidla časová řada očištěná od kalendářních variací).
   - **Časovou řadu odvozených charakteristik** – vznikají na základě absolutních údajů, např. časové řady součtové (např. časová řada klouzavých ročních úhrnů)

Úvod do časových řad

- **Klouzavým ročním úhrnem** rozumíme hodnotu intervalového ukazatele za celé roční období, které končí sledovaným měsícem.

Úvod do časových řad

Měsíc	Počet nehod		Rozdíl roku	Klouzavé roční úhrny
	2005	2006	2006 - 2005	
Leden	16 961	17 219	258	199 262 + 258 = 199 520
Únor	16 375	16 789	414	199 520 + 414 = 199 934
Březen	15 527	17 748	2 221	199 934 + 2 221 = 202 155
Duben	14 168	15 598	1 430	202 155 + 1 430 = 203 585
Květen	16 827	17 031	204	203 585 + 204 = 203 789
Červen	16 707	17 996	1 289	203 789 + 1 289 = 205 078
Červenec	15 937	11 746	-4 191	205 078 - 4 191 = 200 887
Srpen	17 065	13 595	-3 470	200 887 - 3 470 = 197 417
Září	16 536	13 854	-2 682	197 417 - 2 682 = 194 735
Říjen	16 721	15 841	-880	194 735 - 880 = 193 855
Listopad	17 693	15 632	-2 061	193 855 - 2 061 = 191 794
Prosinec	18 745	14 916	-3 829	191 794 - 3 829 = 187 965
∑	199 262	187 965		

Úvod do časových řad

4) Podle způsobu vyjádření údajů rozlišujeme časové řady:
   - **Naturálních ukazatelů** (hodnoty příslušného ukazatele jsou vyjádřeny naturálním kritériem).
   - **Peněžních ukazatelů** (hodnoty ukazatele jsou vyjádřeny v peněžní formě).

Úvod do časových řad

- Za základní princip modelu časové řady se používá jednorozměrný model:

  $y_t = f(t, \varepsilon_t)$,

  kde y_t je hodnota ukazatele v čase t, kde $t = 1,2,\ldots,n$ a ε_t je hodnota náhodné složky v čase t.

Úvod do časových řad

- K tomuto modelu lze přistupovat více způsoby, zpravidla se užívá **klasický (formální) model**, který dekomponuje časovou řadu na složku:
  - Trendovou (T_t).
  - Sezónní (S_t).
  - Cyklickou (C_t).
  - Náhodnou (ε_t).

Úvod do časových řad

- Vlastní rozklad časové řady v aditivním tvaru potom vypadá:

  $y_t = T_t + S_t + C_t + \varepsilon_t = Y_t + \varepsilon_t$,

  kde Y_t se nazývá teoretická (deterministická) složka.

- **Trendem** rozumíme hlavní tendenci dlouhodobého vývoje sledovaného ukazatele v čase – rostoucí trend, klesající trend, řada bez trendu.

Úvod do časových řad

- **Sezónní složka** je pravidelně se opakující odchylka od trendové složky vyskytující se u časových řad s periodicitou menší než 1 rok.

- **Cyklickou složkou** rozumíme kolísání okolo trendu v důsledku dlouhodobého cyklického vývoje s délkou vlny delší než 1 rok.

Úvod do časových řad

- **Náhodná složka** je složka, kterou nelze popsat žádnou funkcí času, jejím zdrojem jsou drobné a nepopsatelné příčiny.

- Nyní nás bude zajímat popis trendové složky pomocí trendových funkcí.

Analýza trendu časových řad

- Nejčastěji se využívají tyto trendové funkce:
  1. Lineární trend.
  2. Parabolický trend.
  3. Exponenciální trend.
  4. Modifikovaný (posunutý) exponenciální trend.
  5. Logistický trend.
  6. Gompertzova křivka.

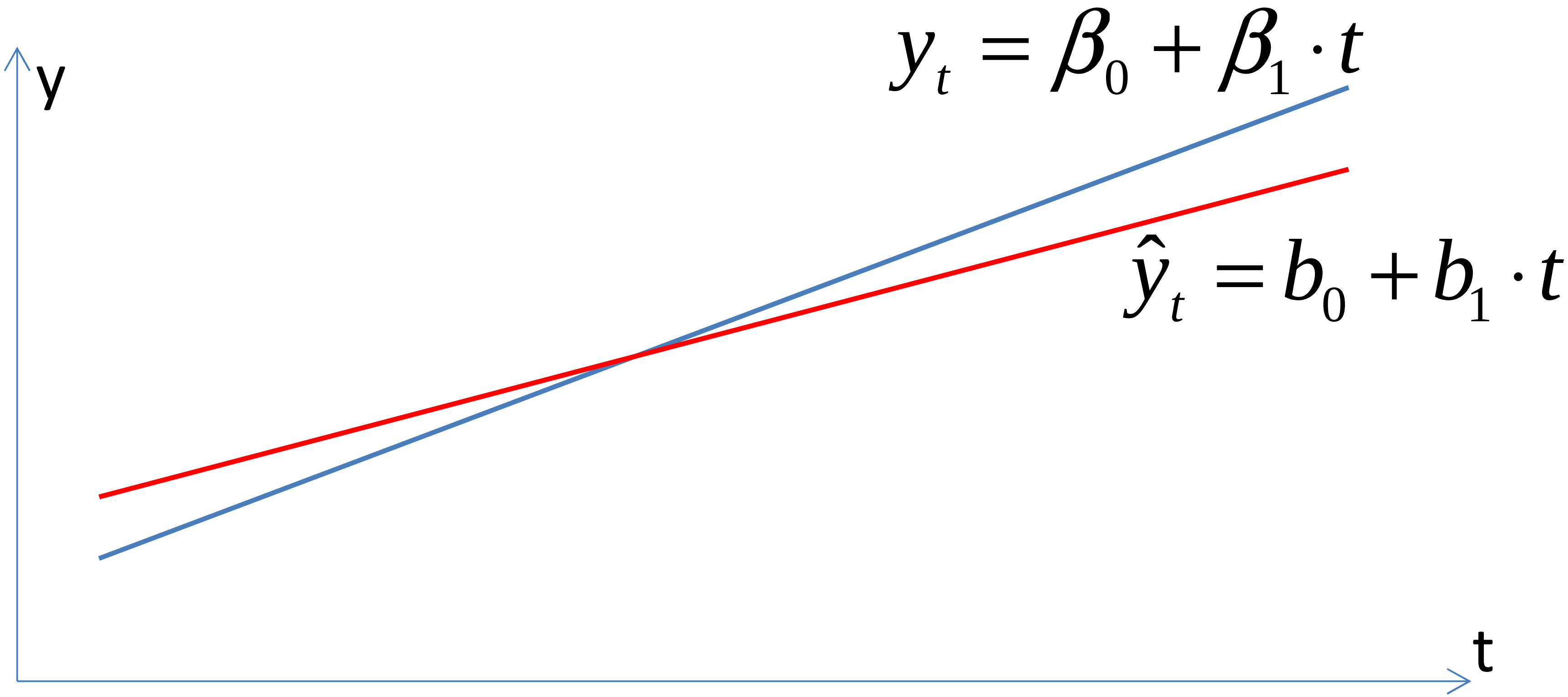
Analýza trendu časových řad

1) Lineární trend

- Mějme hodnoty sledovaného ukazatele y_t pro $t = 1,2,\ldots,n$.
- Skutečný průběh $y_t = \beta_0 + \beta_1 \cdot t$ neznáme, provádíme pouze odhad tohoto trendu ve tvaru:

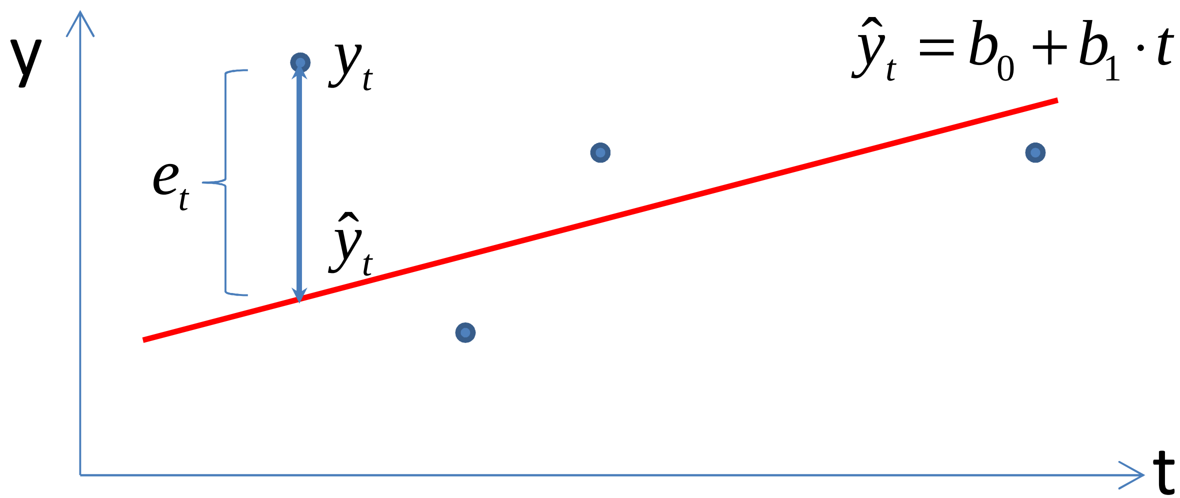
$$\hat{y}_t = b_0 + b_1 \cdot t.$$

Analýza trendu časových řad



Analýza trendu časových řad

- Reziduum (chyba predikce) – $e_t = y_t - \hat{y}_t$ – odchylka naměřené hodnoty od hodnoty předpovídané trendem.



Analýza trendu časových řad

- Úkolem je najít vyrovnávací křivku $\hat{y}_t = b_0 + b_1 \cdot t$ takovou, abychom získali co nejméně rozptýlený soubor reziduí. Můžeme tedy minimalizovat:
  - Součet reziduí $\sum_{t=1}^{n} (y_t - \hat{y}_t)$.
  - Součet absolutních odchylek reziduí $\sum_{t=1}^{n} |y_t - \hat{y}_t|$.
  - Součet druhých mocnin reziduí $\sum_{t=1}^{n} (y_t - \hat{y}_t)^2$.

Analýza trendu časových řad

- Pro odhad parametrů lineárního trendu lze tedy použít metodu nejmenších čtverců, tedy:

$$\varphi = \sum_{t=1}^{n} \left(y_t - \hat{y}_t \right)^2 = \sum_{t=1}^{n} \left(y_t - b_0 - b_1 \cdot t \right)^2 \rightarrow \min .$$

Analýza trendu časových řad

- Jelikož hledáme minimum funkce φ s proměnnými b_0 a b_1, položíme parciální derivace funkce φ rovny nule.

$$\frac{d\varphi}{db_0} = (-2) \cdot \sum_{t=1}^{n} [y_t - b_0 - b_1 \cdot t] = 0,$$

$$\frac{d\varphi}{db_1} = (-2) \cdot \sum_{t=1}^{n} [y_t - b_0 - b_1 \cdot t] \cdot t = 0.$$

Analýza trendu časových řad

- Z první rovnice vyjádříme:

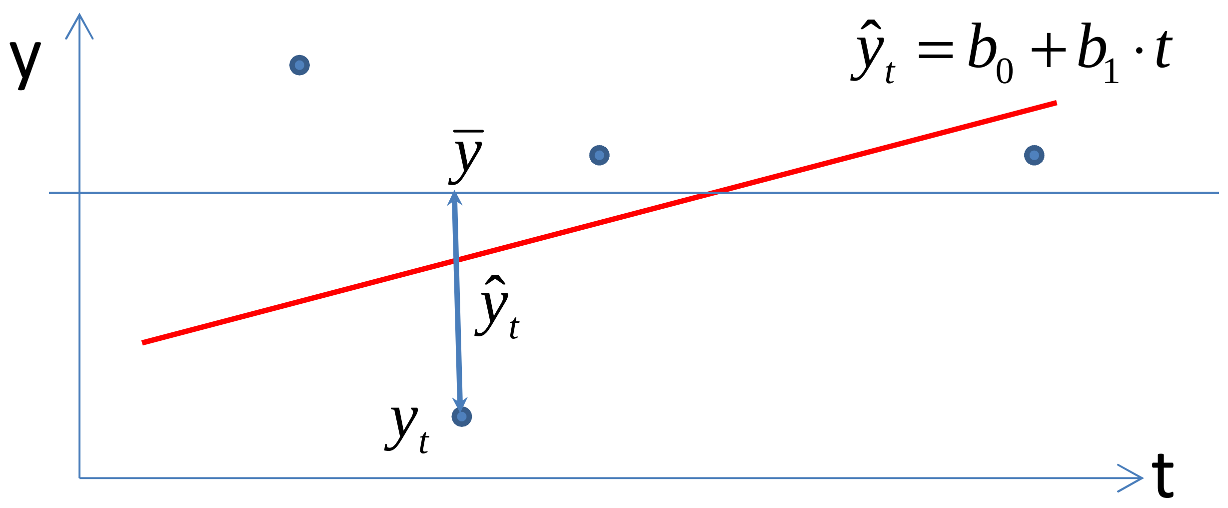
$$b_0 = \frac{\sum_{t=1}^{n} y_t}{n} - b_1 \cdot \frac{\sum_{t=1}^{n} t}{n} = \bar{y} - b_1 \cdot \bar{t}\,.$$

- Dosazením do druhé rovnice a algebraickými úpravami dostaneme:

$$b_1 = \frac{\sum_{t=1}^{n} t \cdot y_t - \bar{y} \cdot \sum_{t=1}^{n} t}{\sum_{t=1}^{n} t^2 - \bar{t} \cdot \sum_{t=1}^{n} t}\,.$$

Analýza trendu časových řad

- Pro účely ověření správnosti zvoleného regresního modelu slouží **index determinace**.



Analýza trendu časových řad

- Označme:
  - Celkový součet čtverců $SS_y = \sum_{i=1}^{n} (y_t - \bar{y})^2$,
  - Součet čtverců modelu $SS_{\hat{y}} = \sum_{i=1}^{n} (\hat{y}_t - \bar{y})^2$,
  - Reziduální součet čtverců $SS_R = \sum_{i=1}^{n} (y_t - \hat{y}_t)^2$.
- Platí:

$$SS_y = SS_{\hat{y}} + SS_R .$$

Analýza trendu časových řad

- Zaveďme $\frac{SS_{\hat{y}}}{SS_y} + \frac{SS_R}{SS_y} = 1$. Je zřejmé, že čím „lepší“ model bude, tím více se bude první zlomek blížit k 1 a naopak.
- Zaveďme index determinace $R^2 = \frac{SS_{\hat{y}}}{SS_y} = \frac{\sum_{t=1}^{n} (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^{n} (y_t - \bar{y})^2}$.
- Index determinace nabývá hodnot z intervalu <0;1>. Velké hodnoty (cca nad 0,8) znamenají, že použitý model se hodí pro popis trendu.

Analýza trendu časových řad

- **Př. 1:** Na základě předchozích sčítání intenzit je známa hodnota RPDI pro předcházející období. Odhadněte rovnici lineárního trendu pro RPDI a extrapolací odhadněte předpokládanou hodnotu RPDI v příštím období.

Analýza trendu časových řad

Rok	t	y_t
1980	1	11523
1985	2	12201
1990	3	12948
1995	4	13578
2000	5	14987
2005	6	16012
2010	7	17065

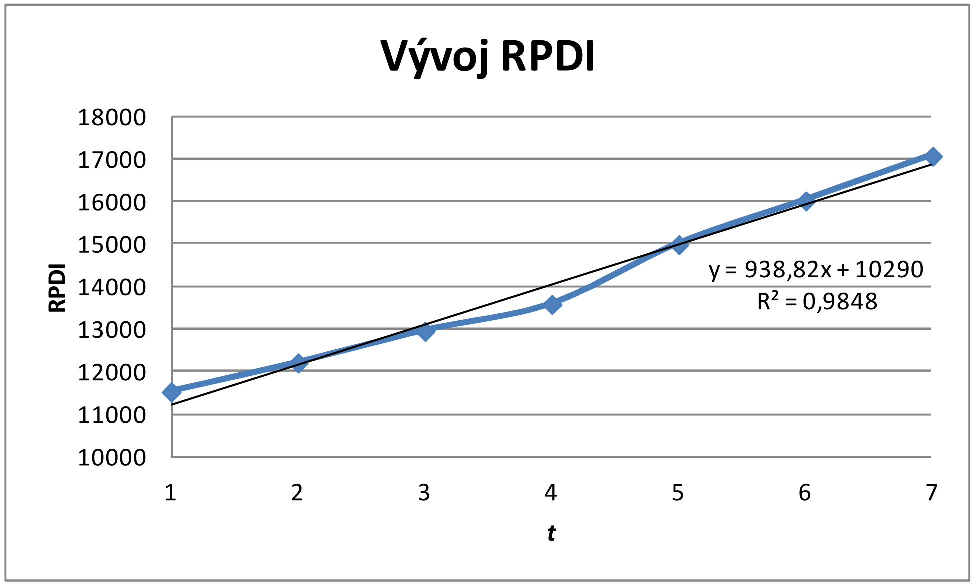
$$b_0 = \bar{y} - b_1 \cdot \bar{t}$$

$$b_1 = \frac{\sum_{t=1}^{n} t \cdot y_t - \bar{y} \cdot \sum_{t=1}^{n} t}{\sum_{t=1}^{n} t^2 - \bar{t} \cdot \sum_{t=1}^{n} t}.$$

$$R^2 = \frac{\sum_{t=1}^{n} \left(\hat{y}_t - \bar{y}\right)^2}{\sum_{t=1}^{n} \left(y_t - \bar{y}\right)^2}$$

Analýza trendu časových řad

Rok	t	y_t	$t \cdot y_t$	t^2	$\hat{y}_t$	$(\hat{y}_t - y_p)^2$	$(y_t - y_p)^2$
1980	1	11523	11523	1	11228	7932471,07	6359763,45
1985	2	12201	24402	4	12167	3525542,70	3399809,16
1990	3	12948	38844	9	13106	881385,67	1203095,59
1995	4	13578	54312	16	14045	0,00	217955,59
2000	5	14987	74935	25	14984	881385,67	887633,16
2005	6	16012	96072	36	15923	3525542,70	3869651,02
2010	7	17065	119455	49	16861	7932471,07	9121262,88
∑	28		419543	140		24678798,89	25059170,86
Průměr	4	14044,86					



b_0	10289,57
b_1	938,82
R^2	0,98

$$\hat{y}_t = 10289{,}57 + 938{,}82 \cdot t$$

Analýza trendu časových řad

- Předpověď pro první následující období (rok 2015) získáme dosazením *t*=8 do rovnice trendu:

$$\hat{y}_8 = 10289{,}57 + 938{,}82 \cdot 8 \doteq 17800 \text{ vozidel} \cdot \text{den}^{-1}.$$

Analýza trendu časových řad

2) Parabolický trend

- Pro odhad průběhu parabolického trendu lze psát:

$$\hat{y}_t = b_0 + b_1 \cdot t + b_2 \cdot t^2 .$$

Analýza trendu časových řad

3) Exponenciální trend

- Pro odhad průběhu trendu lze psát:

$$\hat{y}_t = b_0 \cdot b_1^t \text{ pro } b_1 > 0.$$

- Tento model není lineární v parametrech, nelze přímo použít metodu nejmenších čtverců.

Analýza trendu časových řad

- Odhad parametrů modelu lze získat metodou linearizující transformace (odhad trendové funkce zlogaritmujeme) a aplikací metody nejmenších čtverců.
- K odhadu parametrů lze použít i jiných metod než je metoda nejmenších čtverců – např. metoda vybraných bodů atd.

Analýza trendu časových řad

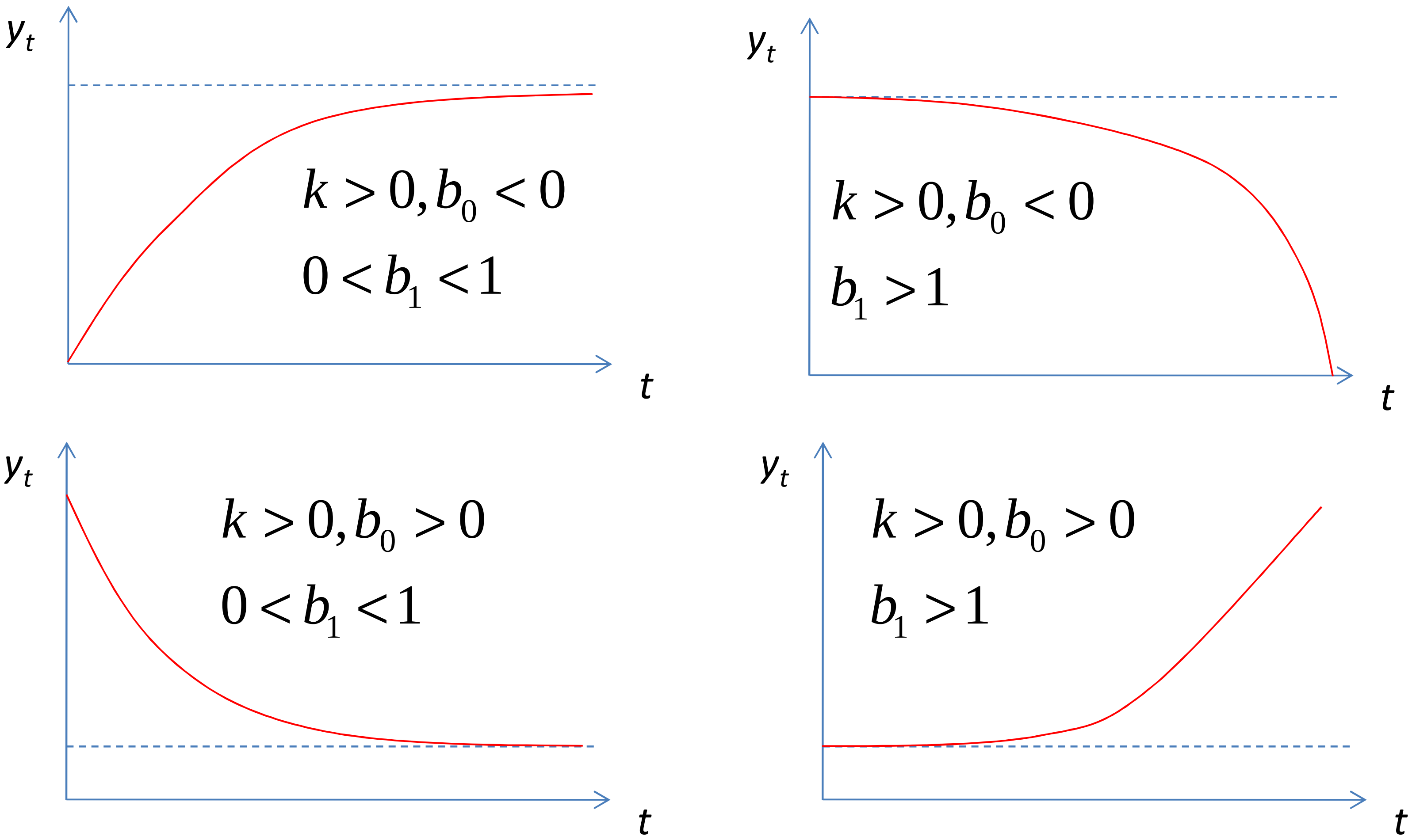
4) Modifikovaný exponenciální trend

- Pro odhad průběhu trendu lze psát:

$\hat{y}_t = k + b_0 \cdot b_1^t \text{ pro } b_1 > 0$.

- Odhad parametrů je již složitější, protože trendovou funkci nemůžeme linearizovat pro použití metody nejmenších čtverců.

Analýza trendu časových řad



Analýza trendu časových řad

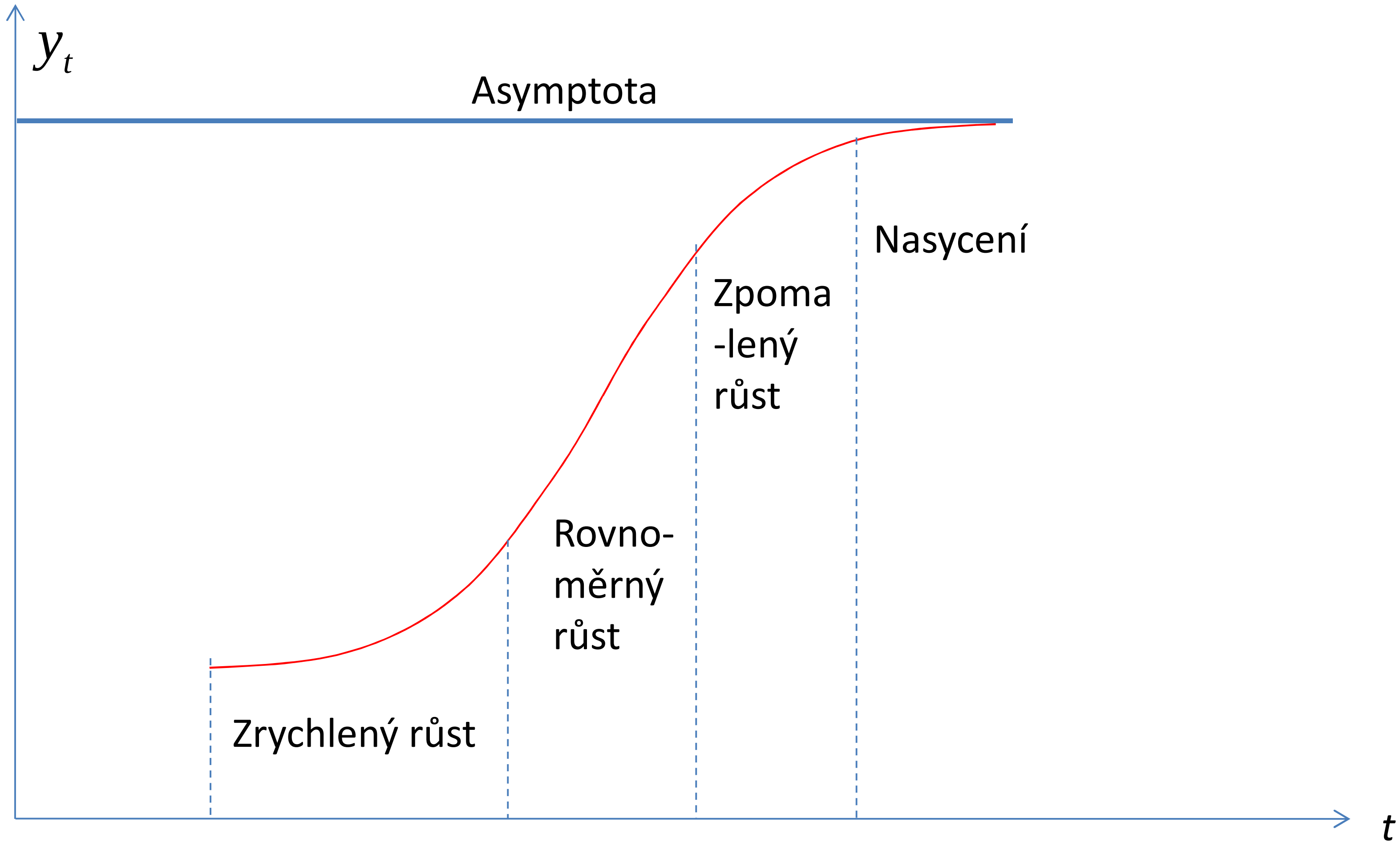
5) Logistický trend

- Pro odhad průběhu trendu lze psát:

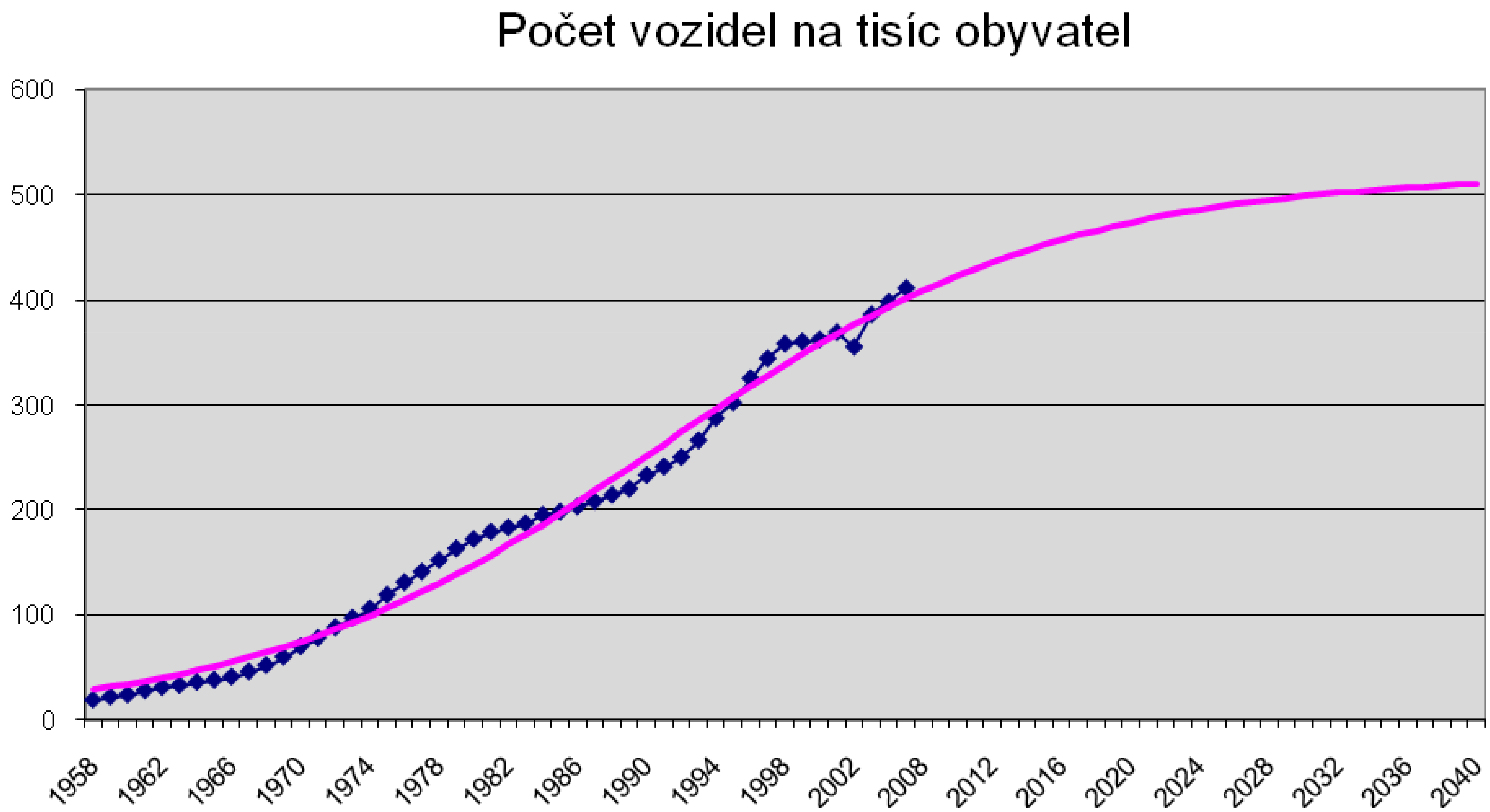
$$\hat{y}_t = \frac{k}{1 + b_0 \cdot b_1^t} \text{ pro } b_0 > 0, b_1 > 0.$$

- Tento funkční předpis je jeden z možných předpisů pro logistický trend.
- Logistická křivka se někdy také nazývá S-křivka.

Analýza trendu časových řad



Analýza trendu časových řad



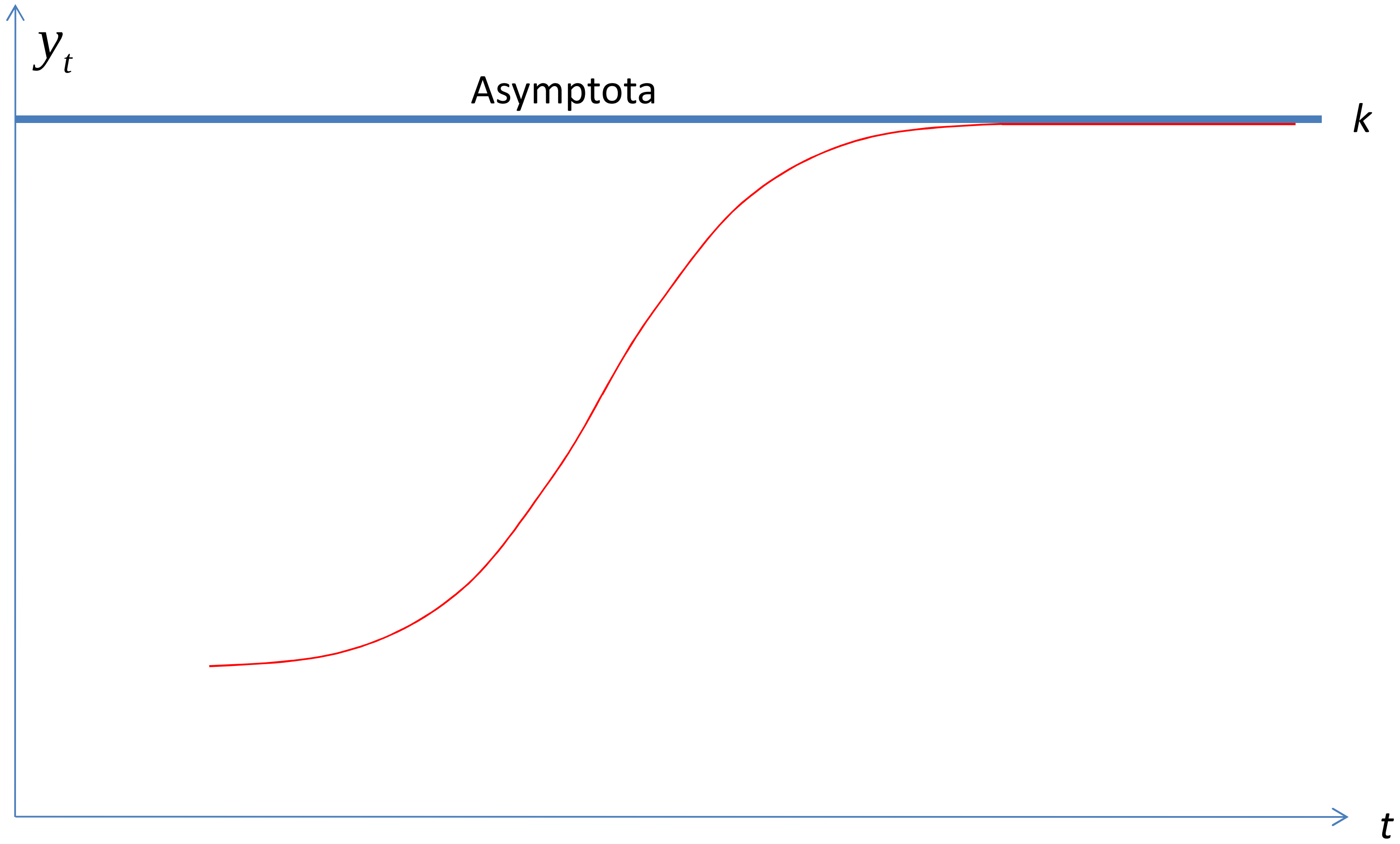
Analýza trendu časových řad

6) Gompertzova křivka

- Má podobný průběh jako logistická křivka, ale není symetrická.
- Pro odhad průběhu trendu lze psát:

$$\hat{y}_t = k \cdot b_0^{b_1^t}.$$

Analýza trendu časových řad



Výběr vhodné trendové funkce

- Nyní se zaměřme na to, na základě jakých kritérií zvolit vhodný typ trendu. Vhodný typ trendu lze volit:
  1) Na základě analýzy grafu studované časové řady (zda jde o rostoucí či klesající trend, zda přichází v úvahu inflexní bod, zda jde o funkci rostoucí do nekonečna nebo rostoucí k nějaké konečné limitě apod.)

Výběr vhodné trendové funkce

2) Dále lze vhodný typ trendové funkce vybrat na základě hodnoty reziduálního součtu čtverců, kdy z možných trendových funkcí vybereme tu s minimálním reziduálním součtem čtverců. Dalším kritériem může být index determinace známý z regresní analýzy, jako vhodný typ trendové funkce vybereme takový, u kterého je index determinace nejvyšší. Snahou je ale použít co nejjednodušší model trendové funkce.

Metoda jednotného součinitele růstu

Metoda jednotného součinitele růstu

- Při prognóze výhledových intenzit dopravních proudů se využívá **metoda jednotného součinitele růstu**.
- *„Silnice se navrhují, případně posuzují na výhledovou padesátirázovou intenzitu v jednom jízdním směru, uvažovanou pro dvacátý rok po uvedení do provozu. Výhledové intenzity nemají překročit návrhové intenzity.“*

Metoda jednotného součinitele růstu

- Výhledovou intenzitu získáme podle vztahu:

$M^v = M^s \cdot K$,

kde: M^v – výhledová intenzita,

M^s – současná intenzita,

K – výhledový koeficient (součinitel růstu).

- Koeficient růstu lze získat např. analýzou časové řady intenzit dopravy.

Metoda jednotného součinitele růstu

- Koeficient růstu je dále možno pro určitou oblast (např. město) odvodit z růstu počtu obyvatel, z počtu vozidel a z růstu jejich proběhu. Pro koeficient růstu můžeme psát:

$$K = \frac{\text{Výhledový počet vozidel}}{\text{Současný počet vozidel}} \times \frac{\text{Výhledový proběh vozidel}}{\text{Současný proběh vozidel}}$$

Metoda jednotného součinitele růstu

- Jelikož se proběh vozidel (počet kilometrů najetých jedním vozidlem za rok) zpravidla nemění, lze psát:

$$K = \frac{\text{Výhledový počet obyvatel}}{\text{Současný počet obyvatel}} \times \frac{\text{Výhledový stupeň automobilizace}}{\text{Současný stupeň automobilizace}}$$

Metoda jednotného součinitele růstu

- Nevýhodou je, že koeficienty jsou jednotné a nezohledňují místní podmínky (nezohledňují např. různé změny v počtech obyvatel v jednotlivých oblastech), proto se používají pouze pro hrubé odhady.

Metoda průměrného součinitele růstu

Metoda průměrného součinitele růstu

- V případech, kdy prognózujeme intenzity dopravy mezi oblastmi, které mají rozdílný koeficient růstu, použijeme **metodu průměrného koeficientu růstu**. Výsledný koeficient růstu mezi dvěma oblastmi bude roven aritmetickému průměru koeficientů růstu obou oblastí.

Metoda průměrného součinitele růstu

- Výhledovou intenzitu stanovíme dle vztahu:

$$M_{ij}^{v} = M_{ij}^{s} \cdot \frac{K_i + K_j}{2},$$

kde M_{ij}^{v} – výhledová intenzita mezi oblastmi *i* a *j*,

M_{ij}^{s} – současná intenzita mezi oblastmi *i* a *j*,

K_i, K_j – koeficienty růstu.

Metoda průměrného součinitele růstu

- Nevýhodou koeficientu růstu je, že nemusí být přímo úměrný k růstu objemu dopravy. Objem dopravy mezi dvěma místy může vzrůst bez změny počtu obyvatel či počtu vozidel, např. na základě vzniku nových pracovních míst apod.
- Proto je vhodnější místo koeficientů růstu stanovovat přímo výhledové objemy dopravy.

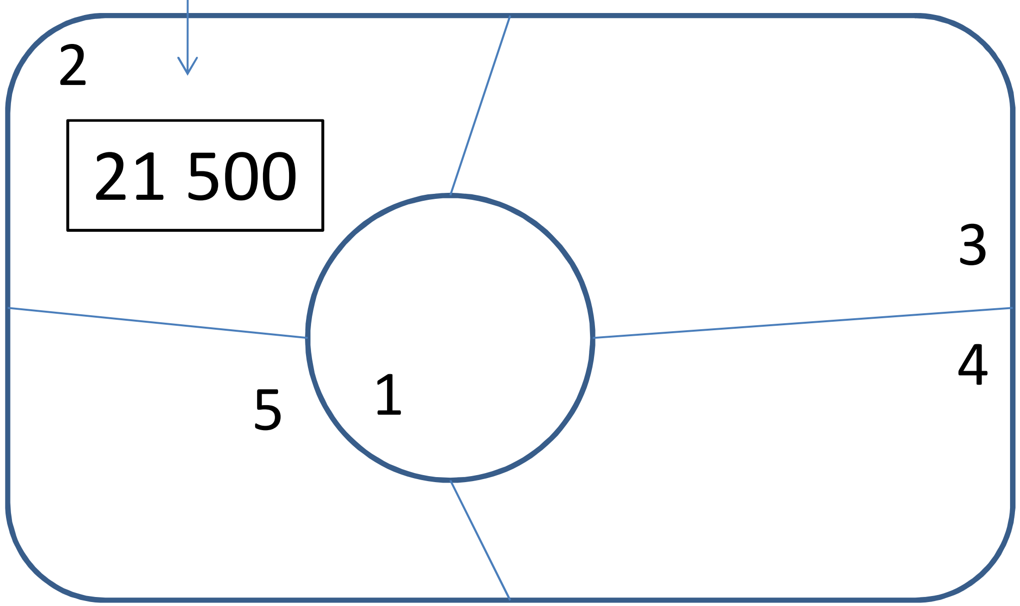
Prognóza dopravy v širším území

Prognóza dopravy v širším území

- Prognóza dopravy v zájmovém území se zpravidla skládá ze 4 fází:
  1. **Určení počtu cest C_i** (výpočet výhledových objemů přepravy) v každé oblasti, na které je území rozděleno. Stanovuje se buď zvlášť počet cest začínajících v oblasti (přepravní produktivita) a počet cest končících v oblasti (přepravní atraktivita) nebo souhrn všech cest majících v oblasti svůj zdroj nebo cíl.

Prognóza dopravy v širším území

Průměrný počet cest za den mající zdroj nebo cíl v oblasti

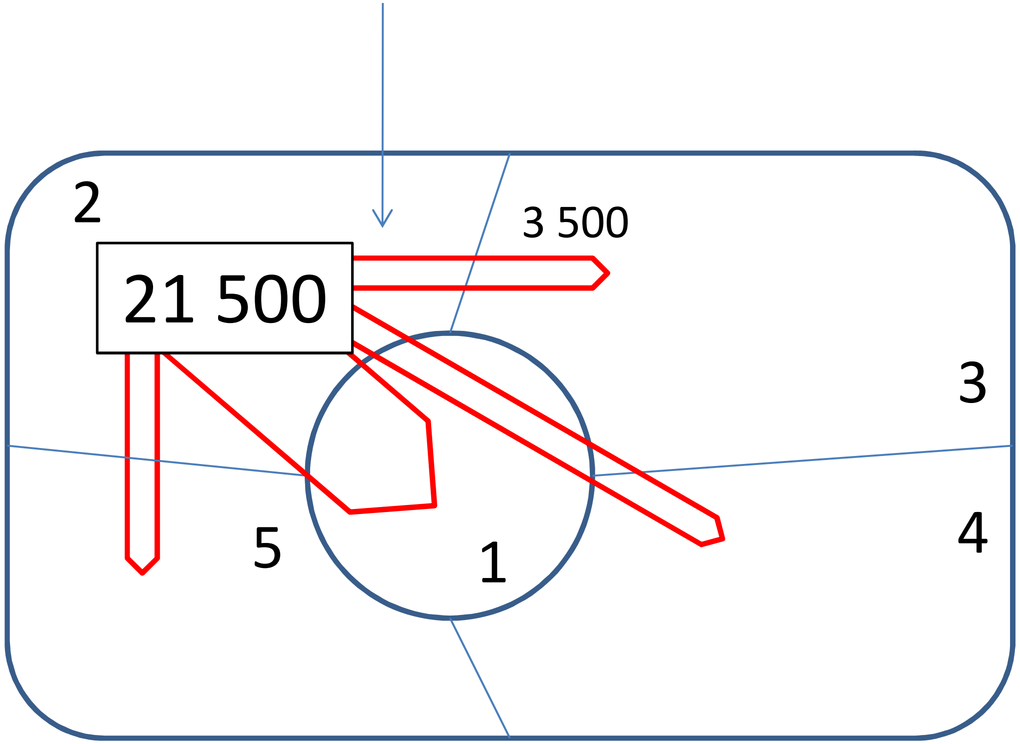


Prognóza dopravy v širším území

2. **Určení mezioblastních vztahů C_{ij}** – rozdělení cest v dané oblasti do přepravních vztahů mezi danou oblastí a ostatními oblastmi (rozdělení přemisťovacích vztahů).

Prognóza dopravy v širším území

Mezioblastní vztah (počet cest za den)

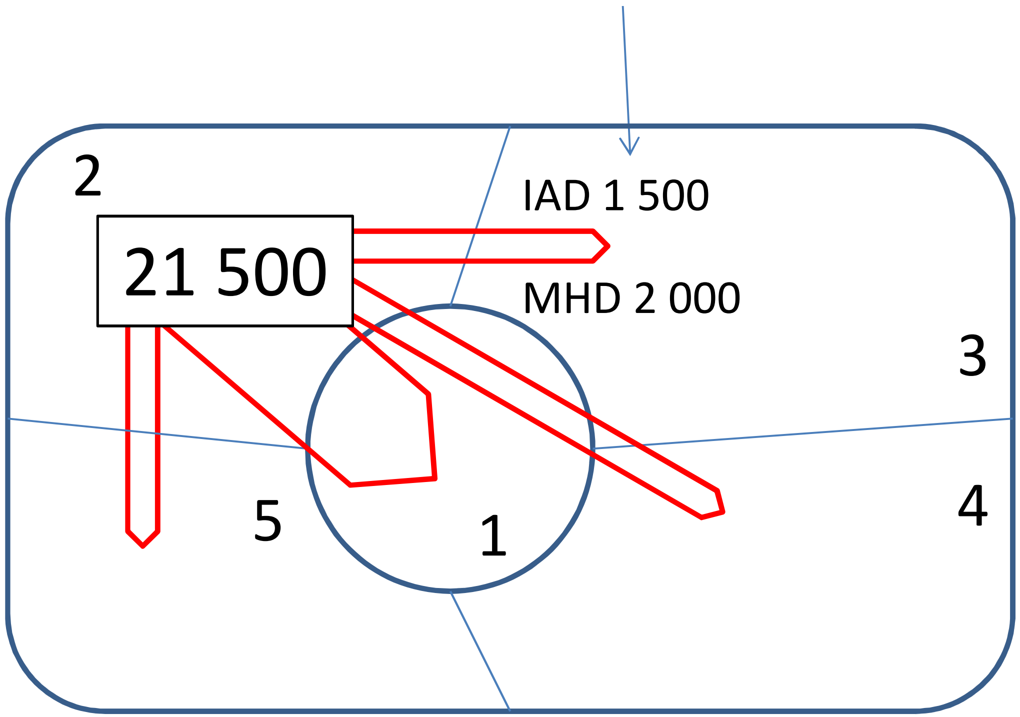


Prognóza dopravy v širším území

3. **Dělba přepravní práce** – stanovení podílů jednotlivých druhů doprav.

Prognóza dopravy v širším území

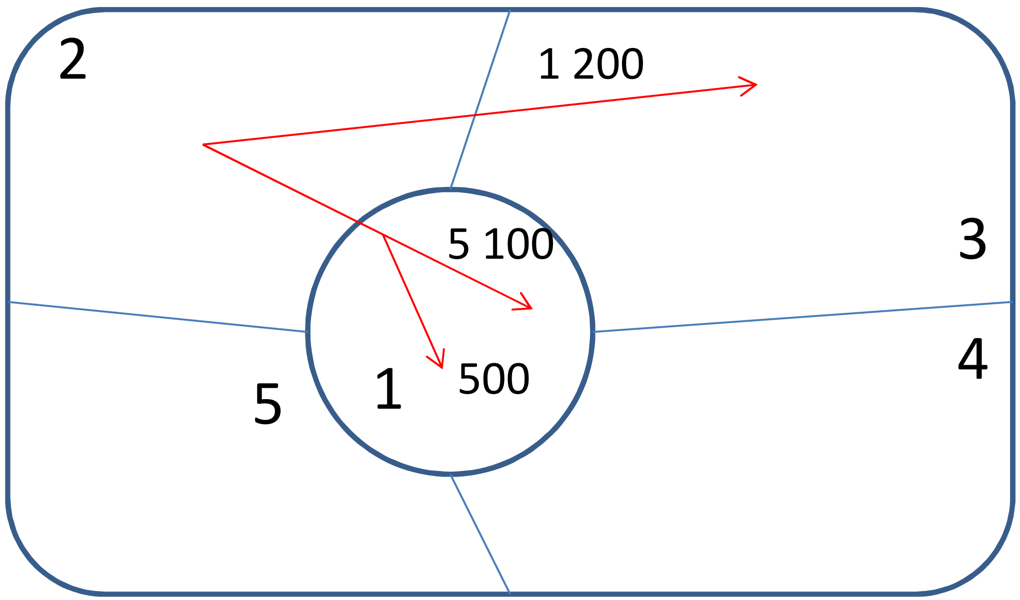
Mezioblastní vztah IAD a MHD (počet cest za den)



Prognóza dopravy v širším území

4. **Určení intenzit** na jednotlivých úsecích sítě (přidělení na síť), výsledkem jsou intenzity na jednotlivých úsecích, resp. křižovatkách apod.

Prognóza dopravy v širším území



Určení výhledových objemů přepravy

Určení výhledových objemů přepravy

- Úkolem je odhadnout objemy přepravy v každé oblasti řešeného území.
- Pokud přepravní vztah v oblasti vzniká, hovoříme o **přepravní produktivitě oblasti**.
- Pokud přepravní vztah do oblasti směřuje, hovoříme o **přepravní atraktivitě oblasti**.

Určení výhledových objemů přepravy

- Každý přepravní vztah je definován:
  - Objektem, který se přepravuje (osoba, náklad).
  - Zdrojem přepravy *i*.
  - Cílem přepravy *j*.
  - Časem *t*, ve kterém je přeprava realizována.
  - Dopravním prostředkem *p*, kterým se přeprava realizuje.
  - Účelem přepravy *u* v osobní přepravě, příp. druhem nákladu v nákladní přepravě.
  - Trasou přepravy *r*.

Určení výhledových objemů přepravy

- Objem přepravy budeme vyjadřovat v počtu cest za časovou jednotku, zpravidla za 1 den.
- Cestou rozumíme jednosměrné přemístění osoby nebo nákladu ze zdrojové oblasti do cílové oblasti a to buď pěšky nebo dopravním prostředkem.

Určení výhledových objemů přepravy

- Objemy přepravy lze stanovit metodami, které můžeme rozdělit do dvou skupin:
  - **Metody regresní a korelační analýzy**.
  - **Metody specifických hybností**.

Určení výhledových objemů přepravy

- Při použití metod regresní a korelační analýzy předpokládáme, že objem dopravy je funkcí jedné nebo více proměnných – počet obyvatel, počet pracovních příležitostí v oblasti atd.
- V další části se tedy podíváme na základní principy regresní a korelační analýzy.

Úvod do regresní a korelační analýzy

Úvod do regresní a korelační analýzy

- K posuzování statistických závislostí slouží regresní a korelační analýza . Úkolem regresní a korelační analýzy je:
  - **Stanovení závislosti** mezi sledovanými kvantitativními znaky (lineární, logaritmická, exponenciální,…), závislost je vyjádřena funkčním předpisem – **regresní analýza**.
  - **Stanovení síly závislosti** mezi sledovanými kvantitativními znaky – **korelační analýza**.

Úvod do regresní a korelační analýzy

- **Vysvětlovaná (závisle) proměnná Y** – proměnná, jejíž chování se snažíme vysvětlit, tedy popsat vyrovnávací křivkou.
- **Vysvětlující (nezávisle) proměnná X** – proměnná, jejíž chování vysvětluje chování závisle proměnné Y. Tato proměnná je příčinnou proměnnou, v důsledku její změny se mění vysvětlovaná proměnná.

Korelační analýza

- Chceme-li posoudit sílu **lineární** závislosti mezi dvěma proměnnými, můžeme použít Pearsonův (výběrový) korelační koeficient:

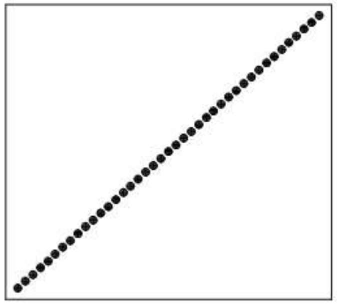
$$r_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

- Pearsonův korelační koeficient nabývá hodnot z intervalu $\langle -1;1 \rangle$.

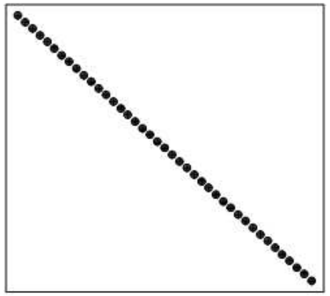
Korelační analýza

- Předpokladem je, že obě náhodné proměnné, pro které počítáme Pearsonův korelační koeficient, pocházejí z normálního rozdělení.
- Hodnotu Pearsonova korelačního koeficientu lze vypočítat v Excelu pomocí funkce PEARSON.

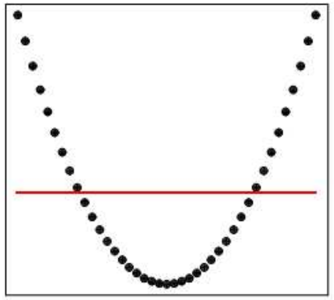
Korelační analýza



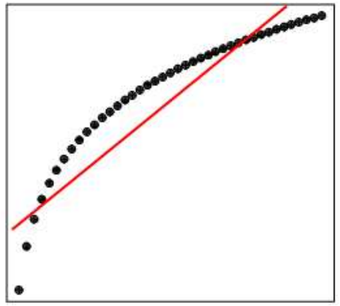
r =1,000



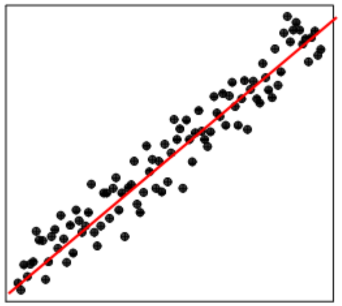
r =-1,000



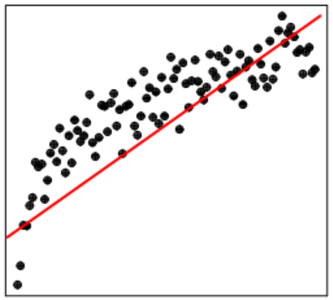
r =0,000



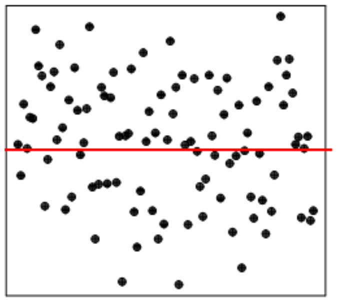
r =0,934



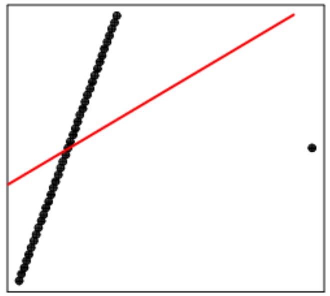
r =0,967



r =0,857



r =-0,143



r =0,608

Regresní analýza

- **Lineární regrese** – závislost proměnných je vyjádřena funkcí lineární v parametrech (resp. se dá na funkci lineární v parametrech převést vhodnou transformací) – např. $y = \beta_0 + \beta_1 \cdot x$.
- **Nelineární regrese** – závislost proměnných je vyjádřena funkcí nelineární v parametrech (a ani nelze na funkci lineární v parametrech převést pomocí žádné transformace).

Regresní analýza

- **Jednoduchá regrese** – studuje závislost jedné proměnné na druhé proměnné.
- **Vícenásobná regrese** – studuje závislost jedné proměnné na několika proměnných.

Jednoduchá regrese

Zajímá nás, zda existuje nějaká závislost mezi průměrným počtem cest v dané oblasti za 1 den a počtem obyvatel v dané oblasti.
Počet obyvatel je v tomto případě vysvětlující proměnná a průměrný počet cest je vysvětlovaná proměnná.

Počet obyvatel [tis. ob.]	Průměrný počet cest [tis. cest/den]
18,1	41,4
83,0	177,7
40,4	96,5
33,8	73,7
29,1	55,0
65,8	150,1
21,1	38,6
99,8	229,5
36,1	96,4
13,2	30,1
81,7	186,4
92,1	229,2
6,4	11,2

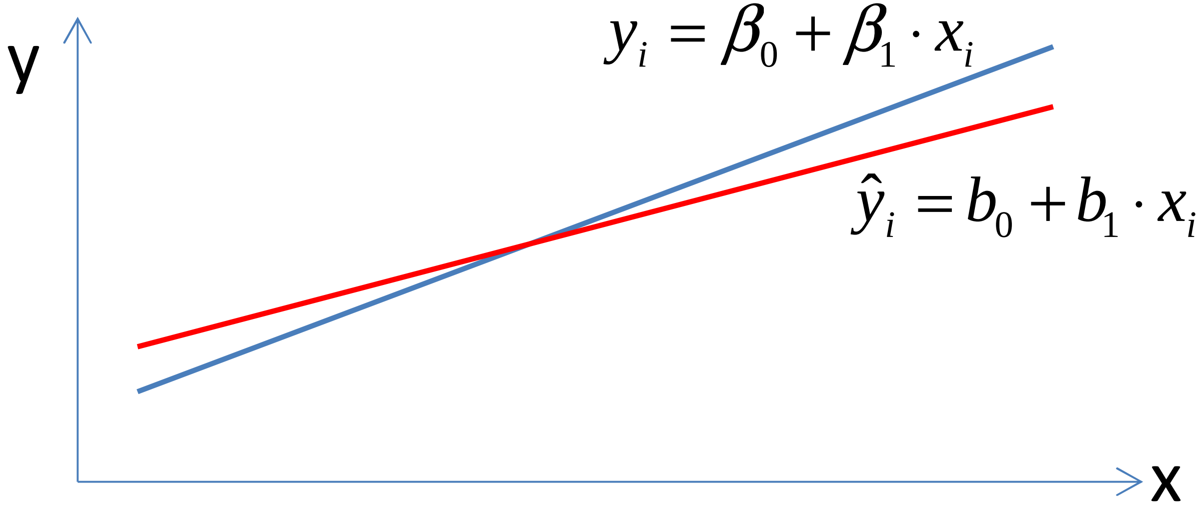
Jednoduchá regrese

- Orientačně („podle oka“) lze druh a sílu závislosti mezi vysvětlující a vysvětlovanou proměnnou posoudit na základě bodového grafu $[x_i, y_i]$ – **korelační pole**.
- Uvažujme nejjednodušší případ a to jednoduchou přímkovou regresi, kterou lze zapsat ve tvaru:

$$y_i = \beta_0 + \beta_1 \cdot x_i \,.$$

Jednoduchá regrese

- Regresní funkce – $y_i = \beta_0 + \beta_1 \cdot x_i$ – skutečná regrese **populace**, v praxi neznámá, proto regresní funkci pouze odhadujeme, zapisujeme tedy $\hat{y}_i = b_0 + b_1 \cdot x_i$.



Jednoduchá regrese

- K nalezení koeficientů vyrovnávací přímky použijeme **metodu nejmenších čtverců**.
- Pro zjednodušení nejdříve upravme vztah pro $\hat{y}_i$ do vhodnější formy – tzv. odchylková forma:

$$\hat{y}_i = b_0 + b_1 \cdot x_1 = \left(b_0 + b_1 \cdot \overline{x}\right) + b_1 \cdot \left(x_1 - \overline{x}\right) = b_0^* + b_1 \cdot \left(x_i - \overline{x}\right).$$

- Potom můžeme psát:

$$\varphi = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n} \left[y_i - b_0^* - b_1 \cdot \left(x_i - \overline{x}\right)\right]^2.$$

Jednoduchá regrese

- Jelikož hledáme minimum funkce φ s proměnnými b_0^* a b_1, položíme parciální derivace funkce φ rovny nule.

$$\frac{d\varphi}{db_0^*} = (-2) \cdot \sum_{i=1}^{n} \left[y_i - b_0^* - b_1 \cdot (x_i - \bar{x}) \right] = 0$$

$$\frac{d\varphi}{db_1} = (-2) \cdot \sum_{i=1}^{n} \left[y_i - b_0^* - b_1 \cdot (x_i - \bar{x}) \right] \cdot (x_i - \bar{x}) = 0$$

Jednoduchá regrese

- Vyřešme nyní první rovnici.

$$(-2)\cdot\sum_{i=1}^{n}\left[y_i - b_0^* - b_1\cdot(x_i-\overline{x})\right]=0$$

$$\sum_{i=1}^{n} y_i - n\cdot b_0^* - b_1\cdot\sum_{i=1}^{n}(x_i-\overline{x})=0$$

$$\sum_{i=1}^{n}(x_i-\overline{x})=0$$

$$n\cdot b_0^* = \sum_{i=1}^{n} y_i$$

$$b_0^* = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$b_0^* = \overline{y} \Rightarrow b_0 + b_1\cdot\overline{x} = \overline{y}$$

$$b_0 = \overline{y} - b_1\cdot\overline{x}$$

Jednoduchá regrese

- Nyní upravme druhou rovnici.

$$(-2)\cdot\sum_{i=1}^{n}\left[y_i - b_0^* - b_1\cdot(x_i-\overline{x})\right]\cdot(x_i-\overline{x})=0$$

$$\sum_{i=1}^{n} y_i\cdot(x_i-\overline{x}) - b_0^*\cdot\sum_{i=1}^{n}(x_i-\overline{x}) - b_1\cdot\sum_{i=1}^{n}(x_i-\overline{x})^2=0$$

$$\sum_{i=1}^{n} y_i\cdot(x_i-\overline{x}) = b_1\cdot\sum_{i=1}^{n}(x_i-\overline{x})^2$$

$$b_1=\frac{\sum_{i=1}^{n} y_i\cdot(x_i-\overline{x})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

$$\sum_{i=1}^{n}(x_i-\overline{x})=0$$

Jednoduchá regrese

- Odvodili jsme tedy vztahy pro koeficienty vyrovnávací přímky ve tvaru:

$$b_1 = \frac{\sum_{i=1}^{n} y_i \cdot (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \text{ a } b_0 = \bar{y} - b_1 \cdot \bar{x} \quad .$$

- Vyrovnávací přímka je potom ve tvaru:

$$\hat{y}_i = b_0 + b_1 \cdot x_i = \bar{y} - b_1 \cdot \bar{x} + b_1 \cdot x_i = \bar{y} + b_1 \cdot (x_i - \bar{x}),$$

prochází tedy vždy bodem $[\bar{x}; \bar{y}]$.

Jednoduchá regrese

x_i	y_i	$x_i - x_p$	$(x_i - x_p) \cdot y_i$	$(x_i - x_p)^2$
18,1	**41,4**	-29,7	-1229,0	881,3
83,0	**177,7**	35,3	6274,1	1246,3
40,4	**96,5**	-7,4	-709,8	54,1
33,8	**73,7**	-13,9	-1026,8	194,0
29,1	**55,0**	-18,6	-1025,4	347,6
65,8	**150,1**	18,1	2712,7	326,7
21,1	**38,6**	-26,6	-1028,8	709,9
99,8	**229,5**	52,0	11945,5	2708,8
36,1	**96,4**	-11,7	-1123,0	135,8
13,2	**30,1**	-34,5	-1037,0	1190,5
81,7	**186,4**	34,0	6335,8	1155,7
92,1	**229,2**	44,3	10155,2	1963,0
6,4	**11,2**	-41,3	-462,7	1706,8
x_p	y_p		Σ	Σ
47,7	108,9		29780,7	12620,7

b_1	b_0
2,36	**-3,76**

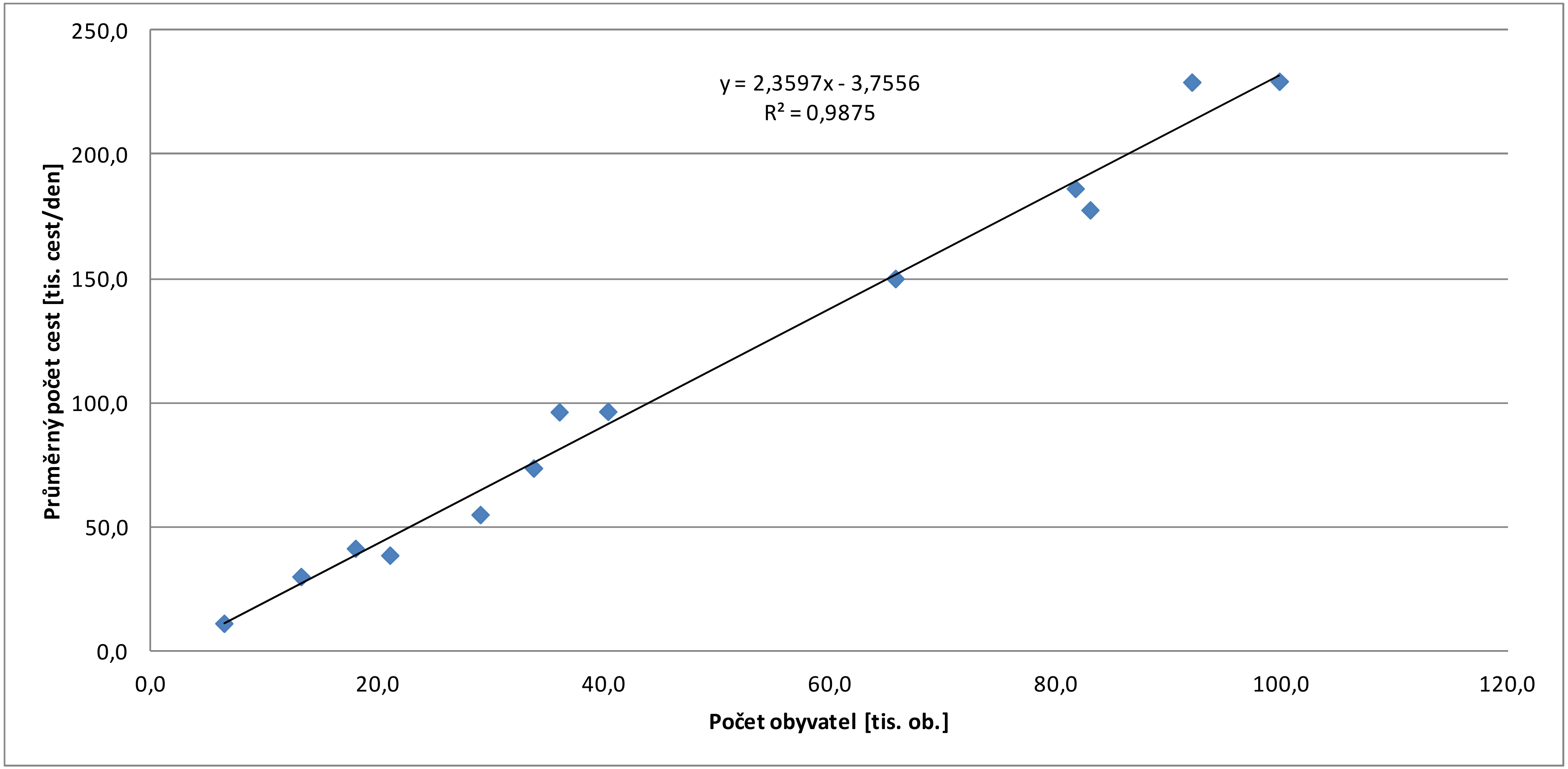
$$\hat{C}_i = -3{,}76 + 2{,}36 \cdot O_i$$

Pozn.

$$x_p = \bar{x}$$

$$y_p = \bar{y}$$

Jednoduchá regrese



$$\hat{C}_i = -3{,}76 + 2{,}36 \cdot O_i$$

Jednoduchá regrese

- Pro účely ověření správnosti zvoleného regresního modelu při použití metody nejmenších čtverců slouží **index determinace**:

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}.$$

- Index determinace nabývá hodnot z intervalu <0;1>. Velké hodnoty (cca nad 0,8) znamenají, že použitý regresní model se hodí pro popis závislosti.

Jednoduchá regrese

$\hat{y}_i$	$(\hat{y}_i - y_p)^2$	$(y_i - y_p)^2$
38,9	4907,2	4556,8
192,2	6939,6	4735,7
91,6	301,0	153,0
76,0	1080,3	1238,2
64,9	1935,4	2905,8
151,6	1819,4	1694,6
46,0	3952,7	4940,9
231,7	15082,8	14547,4
81,4	756,2	157,4
27,5	6628,9	6217,3
189,1	6435,2	6000,5
213,5	10930,3	14472,1
11,4	9503,8	9546,1
	Σ	Σ
	70272,7	71165,7

R^2
0,98745

Jednoduchá regrese

- V případě přímkové regrese platí mezi indexem determinace R^2 a koeficientem korelace $r_{x,y}$ následující vztah:

$$R^2 = r_{x,y}^2 .$$

Jednoduchá regrese

- My jsme se zatím zabývali pouze případem, kdy vyrovnávací křivkou byla přímka. V praxi se používají i jiné regresní modely:

1) **Parabolická regrese**:

$$y_i = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot x_i^2 .$$

2) **Polynomická regrese *n*-tého stupně**:

$$y_i = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot x_i^2 + \ldots + \beta_n \cdot x_i^n .$$

3) **Hyperbolická regrese**:

$$y_i = \beta_0 + \frac{\beta_1}{x_i} .$$

Jednoduchá regrese

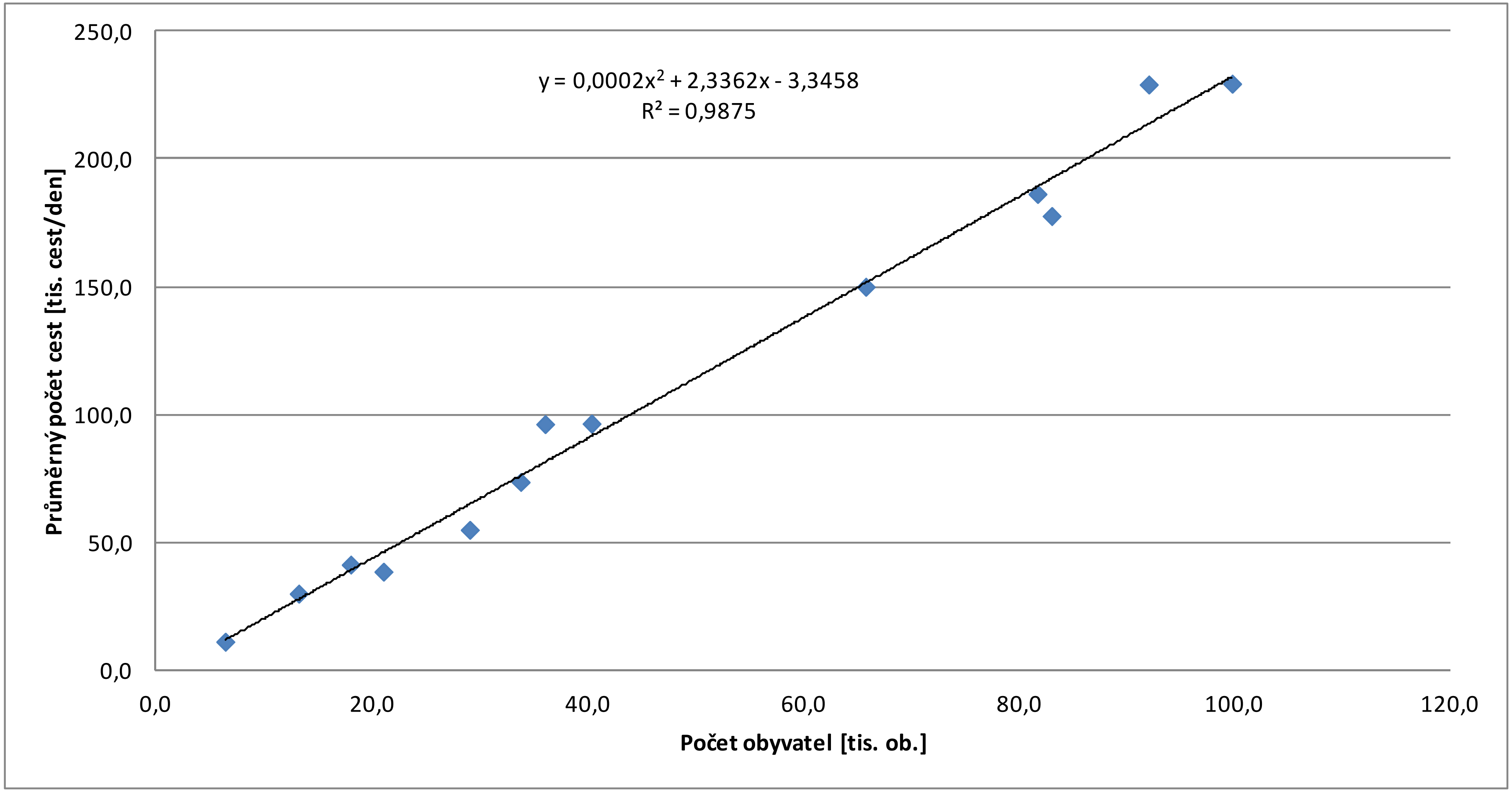
4) **Logaritmická regrese**:

$$y_i = \beta_0 + \beta_1 \cdot \log x_i .$$

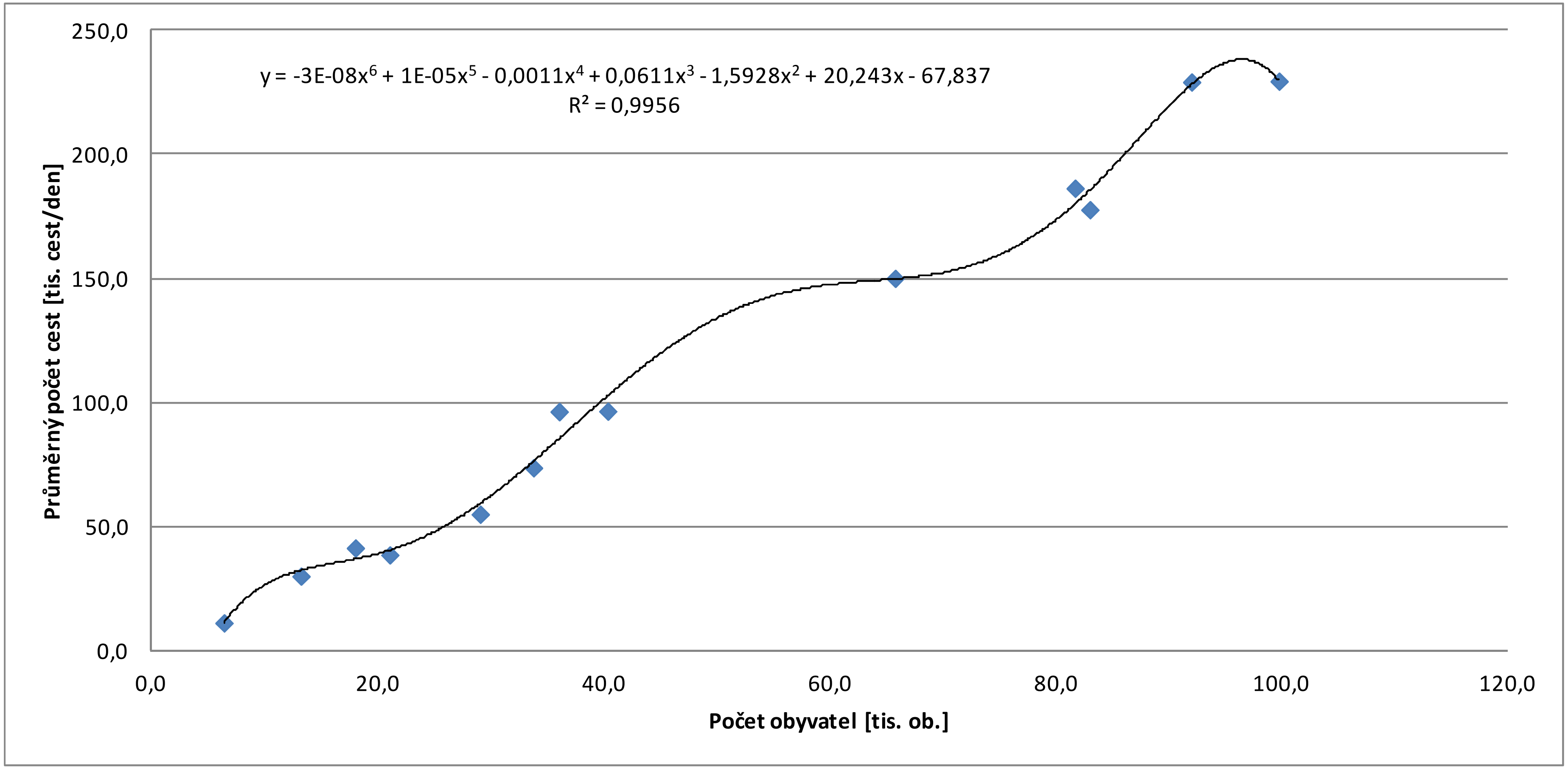
5) **Exponenciální regrese**:

$$y_i = \beta_0 \cdot \beta_1^{x_i} .$$

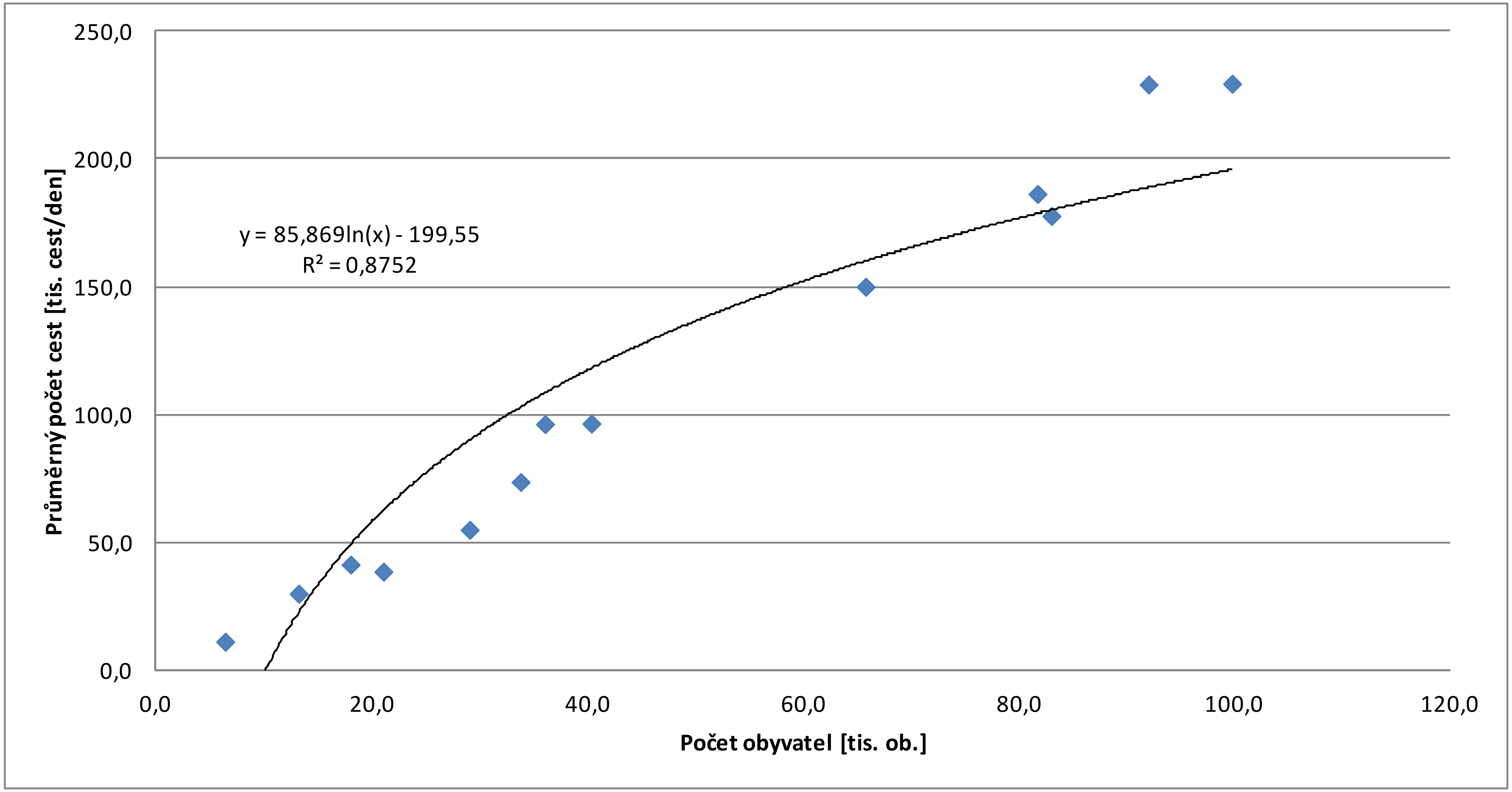
Jednoduchá regrese



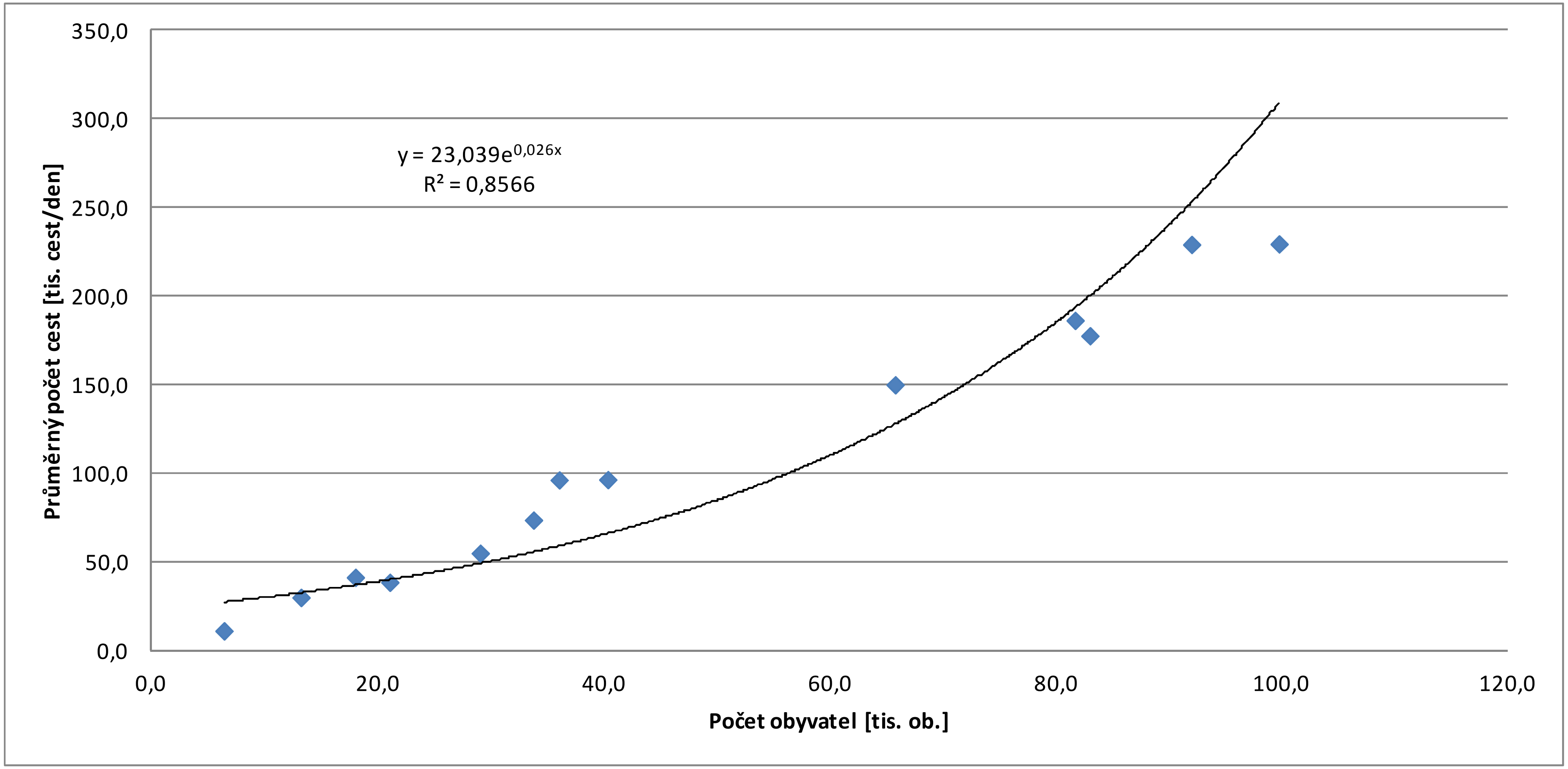
Jednoduchá regrese



Jednoduchá regrese



Jednoduchá regrese



Vícenásobná regrese

- Doposud jsme se zabývali vystižením závislosti vysvětlované proměnné na jedné vysvětlující proměnné, tedy jednoduchou regresí.
- Podívejme se nyní na vícenásobnou regresi, vysvětlovaná proměnná *y* závisí na několika vysvětlujících proměnných $x_1, x_2, \ldots, x_k$.

Vícenásobná regrese

- Uvažujme, že závisle proměnná *y* je lineárně závislá na každé nezávislé proměnné $x_1, x_2, \ldots, x_k$.
- Nechť jsou nezávisle proměnné vzájemně nezávislé.
- Skutečnou regresní funkci pro vícenásobnou přímkovou regresi můžeme tedy vyjádřit ve tvaru:

$$y_i = \beta_0 + \beta_1 \cdot x_{1,i} + \beta_2 \cdot x_{2,i} + \ldots + \beta_k \cdot x_{k,i} \,.$$

Vícenásobná regrese

- Odhad regresní funkce potom můžeme zapsat ve tvaru:

  $$\hat{y}_i = b_0 + b_1 \cdot x_{1,i} + b_2 \cdot x_{2,i} + \ldots + b_k \cdot x_{k,i}.$$

  kde parametry b_1, b_2,…, b_k se nazývají **dílčí regresní koeficienty** a udávají, jak se průměrně změní závisle proměnná při jednotkové změně příslušné nezávislé proměnné.

Vícenásobná regrese

- Jelikož se jedná o model lineární v parametrech, lze použít metodu nejmenších čtverců, tedy:

$$\varphi = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 \cdot x_{1,i} - b_2 \cdot x_{2,i} - \ldots - b_k \cdot x_{k,i})^2 \to \min,$$

$$\frac{d\varphi}{db_0} = 2 \cdot \sum_{i=1}^{n} (y_i - b_0 - b_1 \cdot x_{1,i} - b_2 \cdot x_{2,i} - \ldots - b_k \cdot x_{k,i})^1 \cdot (-1) = 0,$$

$$\frac{d\varphi}{db_1} = 2 \cdot \sum_{i=1}^{n} (y_i - b_0 - b_1 \cdot x_{1,i} - b_2 \cdot x_{2,i} - \ldots - b_k \cdot x_{k,i})^1 \cdot (-x_{1,i}) = 0,$$

$$\frac{d\varphi}{db_2} = 2 \cdot \sum_{i=1}^{n} (y_i - b_0 - b_1 \cdot x_{1,i} - b_2 \cdot x_{2,i} - \ldots - b_k \cdot x_{k,i})^1 \cdot (-x_{2,i}) = 0,$$

$$\vdots$$

$$\frac{d\varphi}{db_n} = 2 \cdot \sum_{i=1}^{n} (y_i - b_0 - b_1 \cdot x_{1,i} - b_2 \cdot x_{2,i} - \ldots - b_k \cdot x_{k,i})^1 \cdot (-x_{k,i}) = 0.$$

Vícenásobná regrese

- Po úpravách získáme:

$$\sum_{i=1}^{n} y_i = n \cdot b_0 + b_1 \cdot \sum_{i=1}^{n} x_{1,i} + b_2 \cdot \sum_{i=1}^{n} x_{2,i} + \ldots + b_k \cdot \sum_{i=1}^{n} x_{k,i} \,,$$

$$\sum_{i=1}^{n} y_i \cdot x_{1,i} = b_0 \cdot \sum_{i=1}^{n} x_{1,i} + b_1 \cdot \sum_{i=1}^{n} x_{1,i}^2 + b_2 \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} + \ldots + b_k \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{k,i} \,,$$

$$\sum_{i=1}^{n} y_i \cdot x_{2,i} = b_0 \cdot \sum_{i=1}^{n} x_{2,i} + b_1 \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} + b_2 \cdot \sum_{i=1}^{n} x_{2,i}^2 + \ldots + b_k \cdot \sum_{i=1}^{n} x_{2,i} \cdot x_{k,i} \,.$$

$$\vdots$$

$$\sum_{i=1}^{n} y_i \cdot x_{k,i} = b_0 \cdot \sum_{i=1}^{n} x_{n,i} + b_1 \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{n,i} + b_2 \cdot \sum_{i=1}^{n} x_{2,i} \cdot x_{n,i} + \ldots + b_k \cdot \sum_{i=1}^{n} x_{n,i}^2 \,.$$

Vícenásobná regrese

- Dostali jsme soustavu k+1 lineárních rovnic s k+1 neznámými parametry regresní funkce.
- Řešením této soustavy dostaneme odhady parametrů hledané regresní funkce.

Vícenásobná regrese

- V případě, že se omezíme na případ, kdy závisle proměnná závisí na dvou nezávisle proměnných, potom dostáváme rovnice ve tvaru:

$$\sum_{i=1}^{n} y_i = n \cdot b_0 + b_1 \cdot \sum_{i=1}^{n} x_{1,i} + b_2 \cdot \sum_{i=1}^{n} x_{2,i},$$

$$\sum_{i=1}^{n} y_i \cdot x_{1,i} = b_0 \cdot \sum_{i=1}^{n} x_{1,i} + b_1 \cdot \sum_{i=1}^{n} x_{1,i}^2 + b_2 \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i},$$

$$\sum_{i=1}^{n} y_i \cdot x_{2,i} = b_0 \cdot \sum_{i=1}^{n} x_{2,i} + b_1 \cdot \sum_{i=1}^{n} x_{1,i} \cdot x_{2,i} + b_2 \cdot \sum_{i=1}^{n} x_{2,i}^2.$$

Vícenásobná regrese

- Vydělíme-li první rovnici *n*, dostáváme po úpravách vztah pro odhad parametru b_0:

$$b_0 = \bar{y} - b_1 \cdot \bar{x}_1 - b_2 \cdot \bar{x}_2 .$$

- Tento vztah můžeme dosadit do zbylých dvou rovnic a vyjádřit tak odhady parametrů b_1 a b_2.

Vícenásobná regrese

- Kvalitu regresního modelu můžeme zase posoudit pomocí indexu determinace, který je definován stejně jako u jednoduché regrese:

$$R^2 = \frac{\sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}.$$

Vícenásobná regrese

- **Př. 5:** Máme k dispozici data o počtu obyvatel, počtu pracovních příležitostí a počtu cest pro jednotlivá území. Úkolem je:
  – Ověřte, zda lze považovat závislosti počtu cest na počtu obyvatel a počtu cest na počtu pracovních příležitostí za statisticky významné (uvažujte lineární závislosti).
  – Nalezněte regresní funkci pro závislost počtu cest na obou proměnných současně.

Vícenásobná regrese

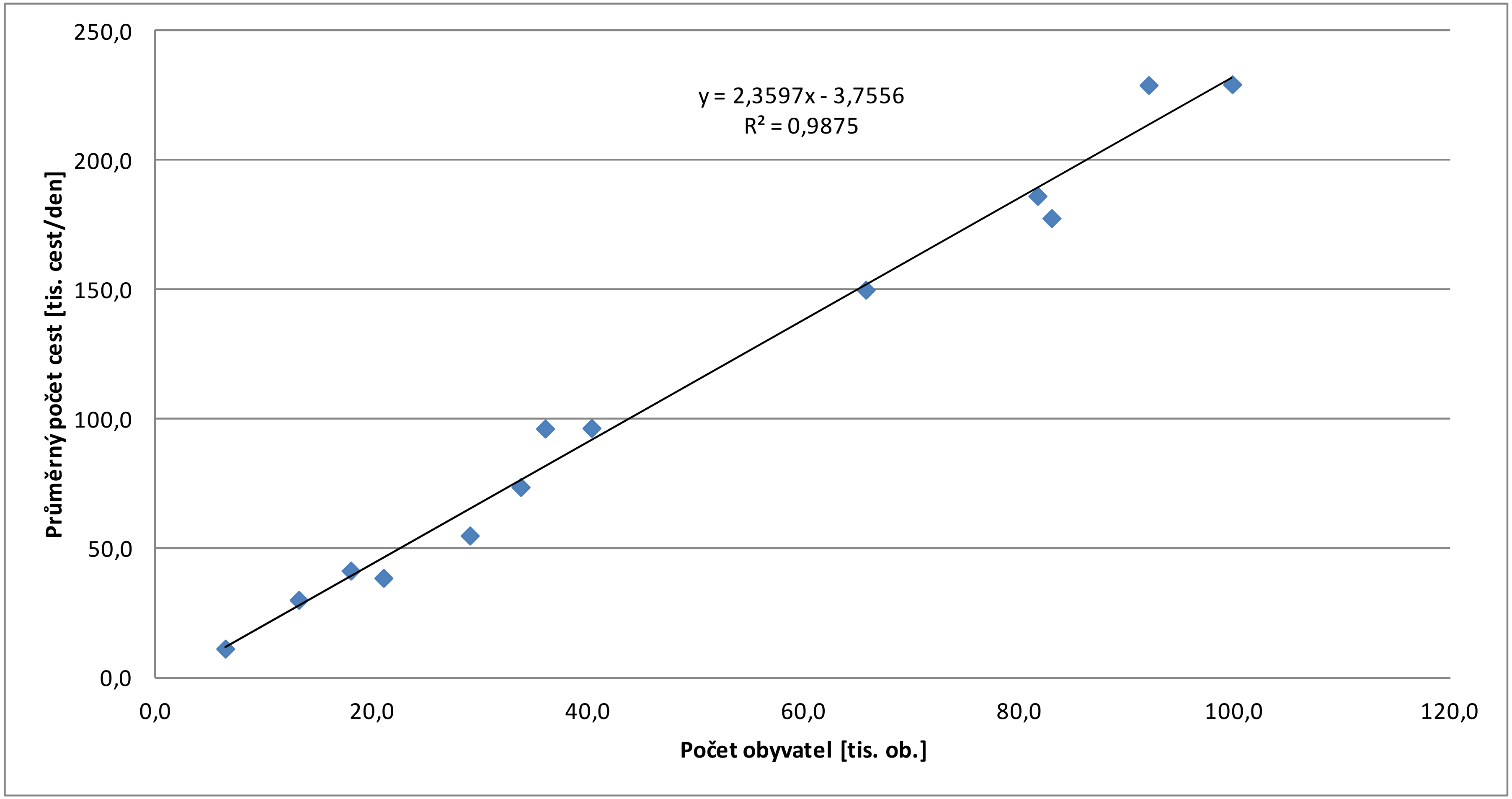
Počet obyvatel [tis. ob.]	Pracovní příležitosti [tis. prac. míst]	Průměrný počet cest [tis. cest/den]
18,1	14,7	41,4
83,0	51,3	177,7
40,4	24,2	96,5
33,8	26,8	73,7
29,1	19,7	55,0
65,8	42,8	150,1
21,1	11,0	38,6
99,8	58,0	229,5
36,1	24,4	96,4
13,2	8,4	30,1
81,7	47,6	186,4
92,1	76,1	229,2
6,4	3,4	11,2

Vícenásobná regrese

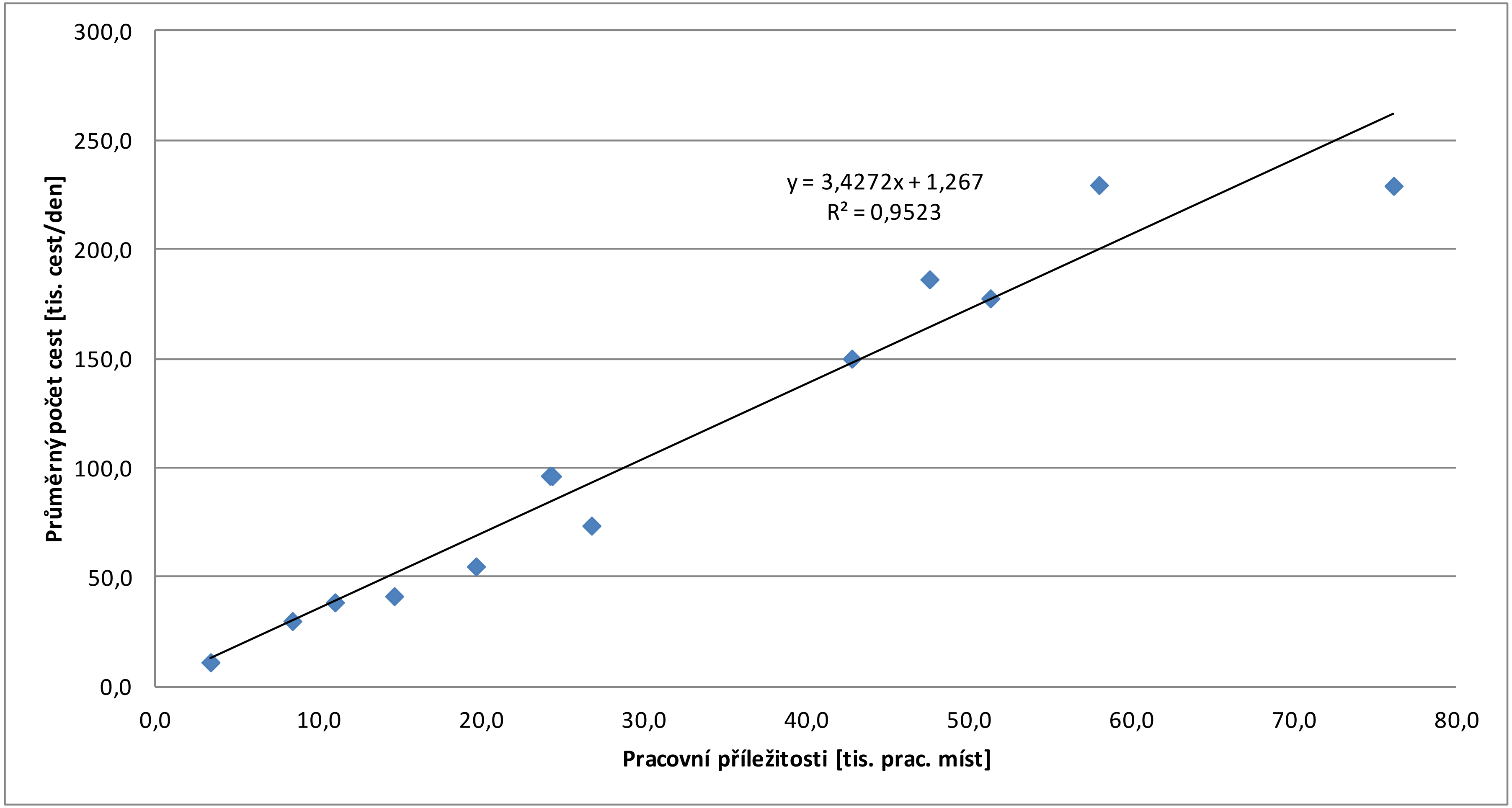
- Nejdříve nalezneme pomocí Excelu obě dílčí regresní funkce a stanovíme hodnotu Pearsonova korelačního koeficientu.

Závislost	Hodnota korelačního koeficientu
Počet cest na počtu obyvatel	0,99371
Počet cest na počtu pracovních příležitostí	0,97585

Vícenásobná regrese



Vícenásobná regrese



Vícenásobná regrese

- Vidíme, že v obou případech máme vysoké hodnoty korelačních koeficientů, lze tedy předpokládat platnost lineárních závislostí.
- Přistoupíme tedy k odhadu koeficientů vícenásobné přímkové regresní funkce.

Vícenásobná regrese

$x_{1,i}$	$x_{2,i}$	y_i	$y_i \cdot x_{1,i}$	$y_i \cdot x_{2,i}$	$x_{1,i} \cdot x_{2,i}$	$(x_{1,i})^2$	$(x_{2,i})^2$
18,1	**14,7**	**41,4**	747,6	606,7	264,6	326,1	214,7
83,0	**51,3**	**177,7**	14759,2	9122,0	4262,6	6896,8	2634,6
40,4	**24,2**	**96,5**	3899,1	2338,8	978,6	1631,4	587,0
33,8	**26,8**	**73,7**	2492,7	1974,1	905,6	1143,5	717,2
29,1	**19,7**	**55,0**	1600,5	1082,5	572,8	846,8	387,4
65,8	**42,8**	**150,1**	9877,6	6422,1	2816,7	4332,3	1831,3
21,1	**11,0**	**38,6**	814,7	424,7	232,1	445,2	121,0
99,8	**58,0**	**229,5**	22903,5	13314,1	5788,7	9958,0	3365,0
36,1	**24,4**	**96,4**	3477,6	2347,7	879,3	1302,5	593,6
13,2	**8,4**	**30,1**	397,9	252,2	111,1	175,3	70,4
81,7	**47,6**	**186,4**	15233,7	8866,2	3888,7	6681,4	2263,3
92,1	**76,1**	**229,2**	21098,3	17449,3	7007,8	8473,2	5795,8
6,4	**3,4**	**11,2**	72,0	37,5	21,5	41,3	11,2
Σ	Σ	Σ	Σ	Σ	Σ	Σ	Σ
620,7	408,3	1415,8	97374,4	64238,1	27730,1	42253,8	18592,6

b_0	b_1	b_2
-3,97	**1,80**	**0,86**

$$\hat{C}_i = -3{,}97 + 1{,}80 \cdot O_i + 0{,}86 \cdot P_i$$

Vícenásobná regrese

y_i	$\hat{y}_i$	$(\hat{y}_i - y_p)^2$	$(y_i - y_p)^2$
38,3	41,1	4584,8	4974,6
177,7	189,5	6516,3	4748,4
96,5	89,5	372,4	150,7
73,7	79,9	838,1	1231,7
55,0	65,3	1894,0	2895,8
150,1	151,2	1797,3	1702,2
38,6	43,4	4272,4	4927,9
229,5	225,4	13593,8	14569,7
96,4	81,9	724,9	155,0
30,1	27,1	6683,2	6202,7
186,4	184,0	5648,0	6014,8
229,2	227,0	13967,8	14494,4
13,1	10,5	9668,9	9157,5
y_p		Σ	Σ
108,8		70562,0	71225,5

R^2
0,99068

Vícenásobná regrese

- Získali jsem odhad regresní závislosti průměrného počtu cest na počtu obyvatel a počtu pracovních příležitostí v dané oblasti.
- Z dosažené hodnoty indexu determinace vidíme, že zvolený model velice dobře vystihuje tuto závislost.
- Pokud bychom chtěli prognózovat výhledové počty cest, dosadili bychom do regresní funkce výhledové počty obyvatel a pracovních příležitostí v dané oblasti.

Určení objemu přepravy - pokračování

Určení objemu přepravy - pokračování

- Chceme-li použít metody regresní a korelační analýzy při stanovení výhledového objemu přepravy, potřebujeme znát:
  - Současné objemy přepravy v každé oblasti (stávající stav).
  - Současné a výhledové hodnoty nezávislých proměnných.
- Výstupem bude buď počet cest celkem nebo výhledové přepravní produktivity a atraktivity jednotlivých oblastí.

Určení objemu přepravy - pokračování

- Prognózu objemu přepravy s využitím regresní a korelační analýzy můžeme rozdělit do následujících kroků:
  1) Výběr nezávisle proměnných a ohodnocení jejich vlivu na závisle proměnnou. Stanovíme tedy korelační koeficienty pro jednotlivé dvojice závisle a nezávisle proměnné a otestujeme, zda je tato závislost statisticky významná. V opačném případě nemá smysl příslušnou nezávisle proměnnou v modelu uvažovat.

Určení objemu přepravy - pokračování

2) Odhadneme koeficienty použité regresní funkce (metoda nejmenších čtverců):

$$\hat{y}_i = b_0 + b_1 \cdot x_{1,i} + b_2 \cdot x_{2,i} + \ldots + b_k \cdot x_{k,i} .$$

3) Známe-li odhad regresní funkce, můžeme výhledový objem přepravy pro danou oblast odhadnout dosazením výhledových hodnot nezávisle proměnných do získané regresní funkce:

$$\hat{y}_i^v = b_0 + b_1 \cdot x_{1,i}^v + b_2 \cdot x_{2,i}^v + \ldots + b_k \cdot x_{k,i}^v .$$

Určení objemu přepravy - pokračování

- Nejčastěji používané nezávisle proměnné při prognóze objemu přepravy C_i se používá počet obyvatel v dané oblasti O_i a počet pracovních příležitostí v dané oblasti P_i.
- Regresní vztah můžeme potom zapsat ve tvaru:

$$\hat{C}_i = b_0 + b_1 \cdot O_i + b_2 \cdot P_i .$$

Určení objemu přepravy - pokračování

- Nevýhodou tohoto přístupu je, nedojde-li ke změně počtu obyvatel a počtu pracovních příležitostí, potom nedojde ani ke změně objemu přepravy v dané oblasti.
- V oblasti ale může dojít i k jiným změnám, které podstatně ovlivní objem přepravy.
- Např. může dojít ke změně demografického složení obyvatelstva a jejich ekonomické aktivity.

Určení objemu přepravy - pokračování

- Dalším problémem může být věcná interpretace parametrů regresní funkce.
- Ve vzorovém příkladě činí odhad parametru b_0=-3,97. Znamená to tedy, že pokud bude oblast bez obyvatel a bez pracovních příležitostí, bude počet cest záporný???

Určení objemu přepravy - pokračování

- Především z těchto důvodů je tento postup nahrazen postupem jiným a to metodami založenými na specifických hybnostech.
- Tyto metody jsou založeny na náhradě jedné rovnice vícenásobné regrese několika rovnicemi jednoduché regrese bez parametru b_0, pomocí kterých odhadneme dílčí počty cest dle jejich účelu. Parametr b_1 je potom označován jako specifická hybnost.
- Sečtením těchto dílčích počtů cest získáme celkový počet cest v dané oblasti.

Určení objemu přepravy - pokračování

- Mezi metody specifických hybností patří:
  - **Metoda specifických hybností obyvatel**.
  - **Metoda specifických hybností domácností**.

Určení objemu přepravy - pokračování

- Metoda specifických hybností obyvatel rozděluje obyvatele dané oblasti na skupiny S_i, např. ekonomicky aktivní obyvatelé atd.
- Pro všechny obyvatele v jedné skupině se předpokládá určitá hodnota specifické hybnosti h_i, která představuje průměrný počet cest vykonané průměrným představitelem dané skupiny za 1 den.

Určení objemu přepravy - pokračování

- Známe-li počet obyvatel O_i dané skupiny, potom pro počet cest C_i pro danou skupinu obyvatel S_i platí vztah:

$$C_i = O_i \cdot h_i \,.$$

- Pro celkový počet cest v dané oblasti potom musí platit:

$$C = \sum_i C_i \,.$$

Určení objemu přepravy - pokračování

- Princip metody specifických hybností domácností je analogický, ale v tomto případě stanovujeme specifické hybnosti pro jednotlivé kategorie domácností.

Určení mezioblastních vztahů

Určení mezioblastních vztahů

- Známe-li počet vznikajících a končících cest v dané oblasti, můžeme přistoupit ke stanovení mezioblastních vztahů, tedy pro každé dvě oblasti určit počet cest mezi nimi. Ke stanovení lze použít:
  - **Analogické metody**.
  - **Syntetické metody**.

Určení mezioblastních vztahů

- Analogické metody jsou založeny na znalostech koeficientů růstu.
- S rozvojem oblastí se bude vyvíjet objem přepravy, přičemž rozvoj oblasti je vyjádřen koeficientem růstu.

Určení mezioblastních vztahů

- Vztah pro výpočet výhledového počtu cest mezi dvěma oblastmi metodami analogickými lze obecně vyjádřit ve tvaru:

  $$C_{i,j}^{v} = C_{i,j}^{s} \cdot f(K),$$

  kde $C_{i,j}^{s}$ představuje současný počet cest mezi oblastmi *i* a *j* a *f*(*K*) je funkce koeficientů růstu oblastí.

Určení mezioblastních vztahů

- Nejpoužívanějšími analogickými metodami jsou:
  - **Metoda jednotného součinitele růstu**.
  - **Metoda průměrného součinitele růstu**.
  - **Metoda detroitská**.
  - **Metoda Fratarova**.
- S prvními dvěma metodami jsme se již seznámili.

Určení mezioblastních vztahů

- **Detroitská metoda** uvažuje, že na počet cest mezi dvěma oblastmi nemá vliv pouze růst těchto oblastí, ale i růst celého města.
- Výhledový počet cest mezi oblastmi *i* a *j* je přímo úměrný současnému počtu cest a součinu koeficientů růstu těchto oblastí a nepřímo úměrný koeficientu růstu celého studovaného území.

Určení mezioblastních vztahů

- Vyjádřeno matematicky:

$$C_{i,j}^{v} = C_{i,j}^{s} \cdot \frac{K_i \cdot K_j}{K},$$

kde koeficient růstu celého území *K* můžeme vyjádřit:

$$K = \frac{\sum_{i=1}^{n} C_i^v}{\sum_{i=1}^{n} C_i^s} = \frac{\sum_{i=1}^{n} C_i^s \cdot K_i}{\sum_{i=1}^{n} C_i^s},$$ kde *n* je celkový počet oblastí daného území.

Určení mezioblastních vztahů

- **Fratarova metoda** předpokládá, že výhledový počet cest mezi dvěma oblastmi *i* a *j* závisí na:
  - Současném počtu cest mezi dvěma oblastmi.
  - Koeficientech růstu obou oblastí.
  - Průměru místních součinitelů obou oblastí. Místní součinitel je vyjádřen poměrem současného počtu cest v oblasti *i* (resp. *j*) a součtu všech současných cest z oblasti *i* (resp. *j*) násobených příslušnými koeficienty růstu.

Určení mezioblastních vztahů

- Zapsáno matematicky:

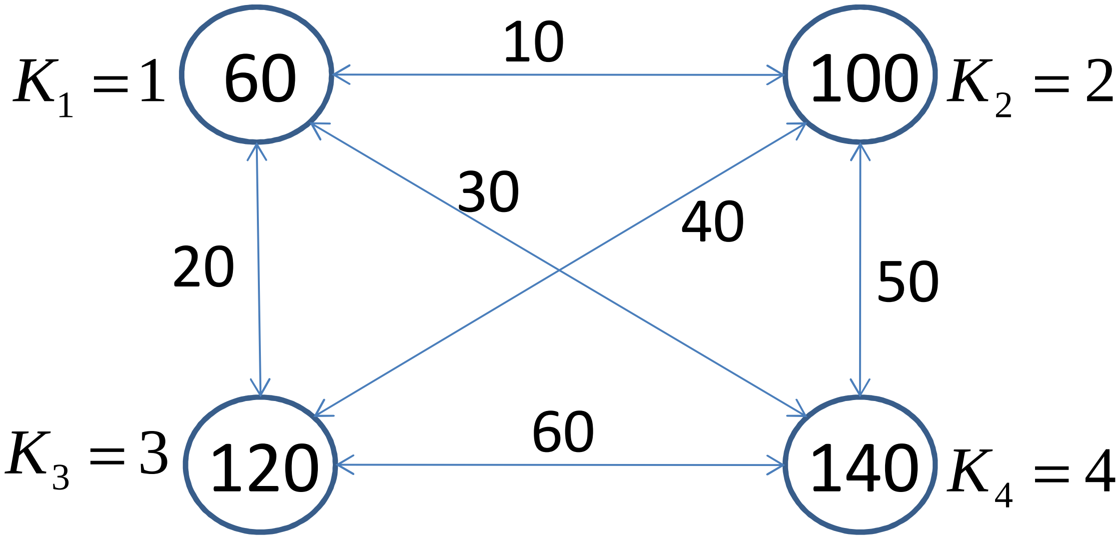
$$C_{i,j}^{v} = C_{i,j}^{s} \cdot K_i \cdot K_j \cdot \frac{L_i + L_j}{2},$$

kde místní součinitel L_i určíme dle vztahu:

$$L_i = \frac{\sum_j C_{i,j}^{s}}{\sum_j C_{i,j}^{s} \cdot K_j},$$ místní součinitel L_j stanovíme analogicky.

Určení mezioblastních vztahů

- **Př. 5:** Vypočítejte výhledové mezioblastní vztahy (tedy počty cest) mezi čtyřmi oblastmi města.



Určení mezioblastních vztahů

- **Metoda průměrných koeficientů růstu:**

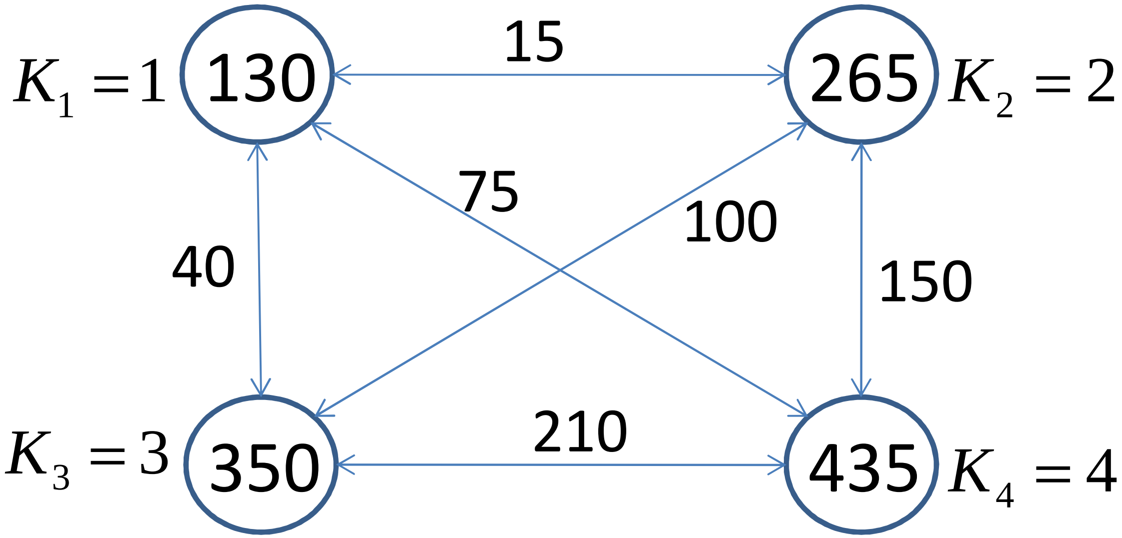
$$C_{i,j}^{v} = C_{i,j}^{s} \cdot \frac{K_i + K_j}{2}$$

Např.

$$C_{1,2}^{v} = C_{1,2}^{s} \cdot \frac{K_1 + K_2}{2} = 10 \cdot \frac{1+2}{2} = 15 \,\text{cest/den},$$

$$C_{1,3}^{v} = C_{1,3}^{s} \cdot \frac{K_1 + K_3}{2} = 20 \cdot \frac{1+4}{2} = 40 \,\text{cest/den atd.}$$

Určení mezioblastních vztahů



Určení mezioblastních vztahů

- **Detroitská metoda:**

$$K = \frac{\sum_{i=1}^{n} C_i^s \cdot K_i}{\sum_{i=1}^{n} C_i^s} = \frac{C_1^s \cdot K_1 + C_2^s \cdot K_2 + C_3^s \cdot K_3 + C_4^s \cdot K_4}{C_1^s + C_2^s + C_3^s + C_4^s} =$$

$$= \frac{60 \cdot 1 + 100 \cdot 2 + 120 \cdot 3 + 140 \cdot 4}{60 + 100 + 120 + 140} = 2{,}81$$

Určení mezioblastních vztahů

$$C_{i,j}^{v} = C_{i,j}^{s} \cdot \frac{K_i \cdot K_j}{K}$$

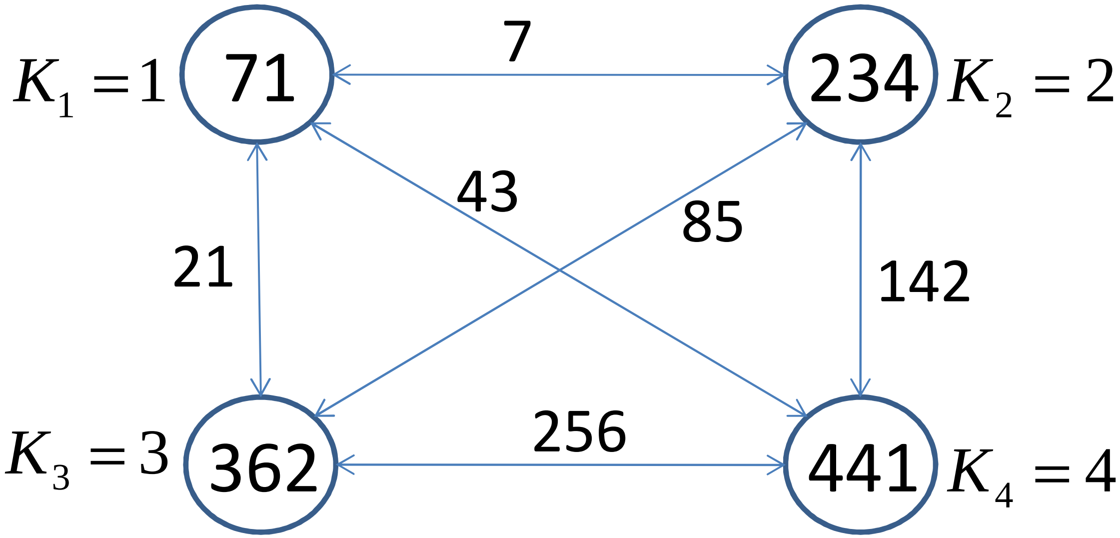
Např.

$$C_{1,2}^{v} = C_{1,2}^{s} \cdot \frac{K_1 \cdot K_2}{K} = 10 \cdot \frac{1 \cdot 2}{2{,}81} \doteq 7\,\text{cest/den},$$

$$C_{1,3}^{v} = C_{1,3}^{s} \cdot \frac{K_1 \cdot K_3}{K} = 20 \cdot \frac{1 \cdot 3}{2{,}81} \doteq 21\,\text{cest/den},$$

$$C_{1,4}^{v} = C_{1,4}^{s} \cdot \frac{K_1 \cdot K_4}{K} = 30 \cdot \frac{1 \cdot 4}{2{,}81} \doteq 43\,\text{cest/den atd.}$$

Určení mezioblastních vztahů



Určení mezioblastních vztahů

- **Fratarova metoda:**

$$L_i = \frac{\sum_j C_{i,j}^s}{\sum_j C_{i,j}^s \cdot K_j}$$

$$L_1 = \frac{C_{1,2}^s + C_{1,3}^s + C_{1,4}^s}{C_{1,2}^s \cdot K_2 + C_{1,3}^s \cdot K_3 + C_{1,4}^s \cdot K_4} = \frac{10+20+30}{10 \cdot 2 + 20 \cdot 3 + 30 \cdot 4} = 0{,}3$$

$$L_2 = \frac{C_{2,1}^s + C_{2,3}^s + C_{2,4}^s}{C_{2,1}^s \cdot K_1 + C_{2,3}^s \cdot K_3 + C_{2,4}^s \cdot K_4} = \frac{10+40+50}{10 \cdot 1 + 40 \cdot 3 + 50 \cdot 4} \doteq 0{,}3$$

Určení mezioblastních vztahů

$$L_3 = \frac{C_{3,1}^s + C_{3,2}^s + C_{3,4}^s}{C_{3,1}^s \cdot K_1 + C_{3,2}^s \cdot K_2 + C_{3,4}^s \cdot K_4} = \frac{20+40+60}{20 \cdot 1 + 40 \cdot 2 + 60 \cdot 4} \doteq 0{,}35$$

$$L_4 = \frac{C_{4,1}^s + C_{4,2}^s + C_{4,3}^s}{C_{4,1}^s \cdot K_1 + C_{4,2}^s \cdot K_2 + C_{4,3}^s \cdot K_3} = \frac{30+50+60}{30 \cdot 1 + 50 \cdot 2 + 60 \cdot 3} \doteq 0{,}45$$

Určení mezioblastních vztahů

$$C_{i,j}^{v} = C_{i,j}^{s} \cdot K_i \cdot K_j \cdot \frac{L_i + L_j}{2}$$

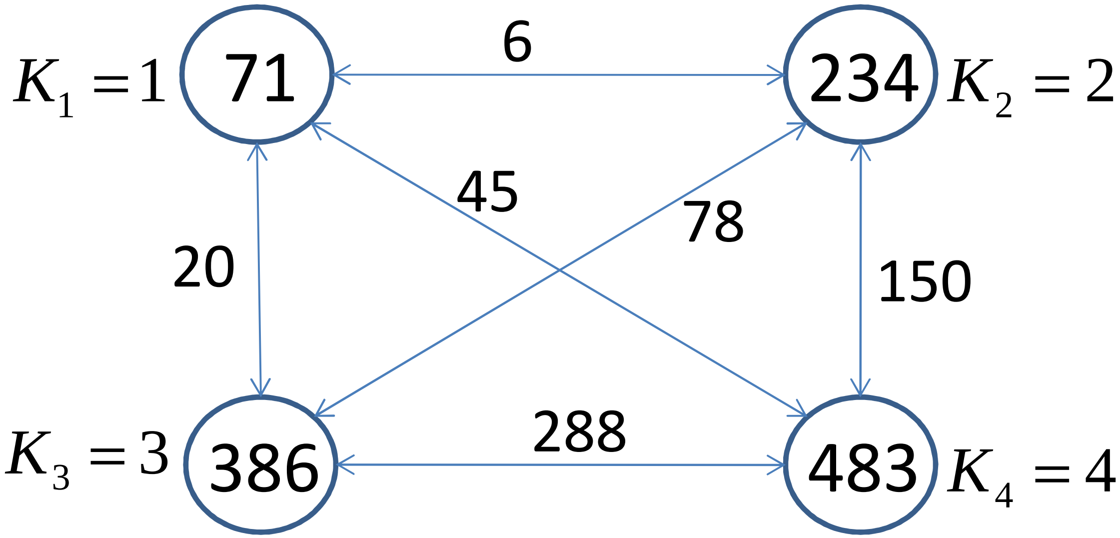
Např.

$$C_{1,2}^{v} = C_{1,2}^{s} \cdot K_1 \cdot K_2 \cdot \frac{L_1 + L_2}{2} = 10 \cdot 1 \cdot 2 \cdot \frac{0{,}3 + 0{,}3}{2} = 6\,\text{cest/den},$$

$$C_{1,3}^{v} = C_{1,3}^{s} \cdot K_1 \cdot K_3 \cdot \frac{L_1 + L_3}{2} = 20 \cdot 1 \cdot 3 \cdot \frac{0{,}3 + 0{,}35}{2} \doteq 20\,\text{cest/den},$$

$$C_{1,4}^{v} = C_{1,4}^{s} \cdot K_1 \cdot K_4 \cdot \frac{L_1 + L_4}{2} = 30 \cdot 1 \cdot 4 \cdot \frac{0{,}3 + 0{,}45}{2} = 45\,\text{cest/den atd.}$$

Určení mezioblastních vztahů



Určení mezioblastních vztahů

Oblast	Očekávaný výhledový objem přepravy $C_i^v=C_i \cdot K_i$	Metoda průměrných koeficientů růstu	Detroitská metoda	Fratarova metoda
1 (K_1=1)	60	130 (+125%)	71 (+18%)	70 (+17%)
2 (K_2=2)	200	265 (+32%)	234 (+17%)	234 (+17%)
3 (K_3=3)	360	350 (-3%)	362 (-1%)	385 (+7%)
4 (K_4=4)	560	435 (-22%)	441 (-21%)	483 (-14%)

Určení mezioblastních vztahů

- Z tabulky vidíme, že vzniká nesoulad mezi námi očekávanými objemy přeprav pro jednotlivé oblasti s objemy přeprav vypočtenými jednotlivými metodami, neplatí vztah $C_i^v \neq \sum_j C_{i,j}^v$.
- Srovnáním dosažených výsledků vidíme, že největší nesoulad vzniká při použití metody průměrných koeficientů růstu (rozdíl až o 125%). Je to dáno tím, že tato metoda neuvažuje vliv růstu ostatních oblastí.

Určení mezioblastních vztahů

- Tento nesoulad je třeba odstranit **balancováním modelu**.
- Balancování modelu se provádí iteračně.
- Při balancování metody průměrných součinitelů růstu postupujeme následujícím postupem.

Určení mezioblastních vztahů

- Výsledky dosažené v rámci př. 5 považujeme za výsledky 1. iterace, tedy:

$$C_{i,j}^{v,1} = C_{i,j}^{s} \cdot \frac{K_i + K_j}{2}.$$

- Na základě tohoto výpočtu stanovíme opravné hodnoty součinitelů růst po 1. iteraci pomocí vztahu:

$$K_i^1 = \frac{C_i^s \cdot K_i}{C_i^{v,1}}, kde\, C_i^{v,1} = \sum_j C_{i,j}^{v,1}.$$

Určení mezioblastních vztahů

- Pomocí těchto opravných součinitelů růstu znovu spočítáme mezioblastní vztahy po 2. iteraci pomocí vztahu:

$$C_{i,j}^{v,2} = C_{i,j}^{v,1} \cdot \frac{K_i^1 + K_j^1}{2}.$$

- Potom spočítáme opravné součinitele růstu po 2. iteraci atd.

Určení mezioblastních vztahů

- Výpočet ukončíme po dosažení požadované přesnosti.
- K ukončení výpočtu dochází při *n*-té iteraci, přiblíží-li se dostatečně opravné součinitele růstu hodnotě 1, zpravidla se uvažuje, že

$$0{,}9 < K_{i,j}^{n} < 1{,}1.$$

- Balancování dalších metod je analogické.

Určení mezioblastních vztahů

- Syntetické metody odvozují výhledové dopravní vztahy ze strukturálních veličin na základě studia vzniku a rozdělování přemísťovacích vztahů. Při popisu těchto vztahů je využíváno analogie se zákony jiných vědních oborů (např. gravitační zákon apod.).
- Mezi syntetické metody patří **Metoda přitažlivosti (Gravitační metoda)**.

Gravitační modely

Gravitační modely

- Gravitační model použijeme, máme-li proveden přepravní průzkum (počty cest v daném období) pouze na vybraných relacích, na základě kterého potom odhadneme počty cest na zbývajících relacích.

Gravitační modely

- Gravitační model předpokládá, že vysoká přepravní poptávka nastává u oblastí s vysokou přepravní produktivitou a atraktivitou nacházející se blízko sebe.
- Se snižující se přepravní produktivitou a atraktivitou a se zvyšující se vzdáleností se přepravní poptávka snižuje.

Gravitační modely

- Gravitační metoda je analogie Newtonova gravitačního zákona. Počet cest mezi oblastmi *i* a *j* lze vyjádřit obecným vztahem:

$$C_{i,j} = k \cdot \frac{C_i \cdot A_j}{f_{i,j}},$$

kde $C_{i,j}$ je počet cest mezi oblastmi *i* a *j*,

k je vhodná konstanta,

Gravitační modely

C_i je přepravní produktivita oblasti *i* (počet cest začínajících v oblasti *i*),

A_j je přepravní atraktivita cílové oblasti *j (*počet cest končících v oblasti j),

$f_{i,j}$ je **odporová funkce**.

Gravitační modely

- Odporová funkce zohledňuje vliv faktorů jako např. vzdálenosti, ceny přepravy či doby přepravy na celkový počet cest mezi dvěma oblastmi *i* a *j*.
- Odporovou funkci můžeme například vyjádřit ve tvaru:

$$f_{i,j} = D_{i,j}^{\alpha},$$

kde α je vhodná konstanta,

$D_{i,j}$ je vzdálenost mezi oblastmi *i*,*j*.

Gravitační modely

- Je zřejmé, že musíme zajistit splnění těchto okrajových podmínek:

$$C_i = \sum_j C_{i,j} \quad \forall i,$$

$$A_j = \sum_i C_{i,j} \quad \forall j.$$

Gravitační modely

- Abychom mohli použít gravitační model, musíme získat hodnoty konstant k a α.
- Tyto konstanty získáme na základě znalostí počtu cest v relacích, u který jsme realizovali dopravní průzkum.
- Tomuto postupu se říká **kalibrace gravitačního modelu**.

Gravitační modely

- Kalibraci modelu lze realizovat např. s využitím doplňku Řešitel v Excelu.
- Pro potřeby kalibrace potřebujeme znát:
  - Přepravní produktivity C_i a atraktivity A_i jednotlivých oblastí.
  - Vzdálenosti (příp. ceny za přepravu, dobu přepravy) $D_{i,j}$ mezi jednotlivými oblastmi.
  - Počty cest $C_{i,j}^{skut}$ mezi oblastmi, které jsme zjistili dopravním průzkumem.

Gravitační modely

- Uvažujme, že dopravním průzkumem jsme zjistili počty cest pro *n* relací. Postup při kalibraci je následující:
  1) Zvolíme počáteční hodnoty parametrů k a α.
  2) Stanovíme odchylky $\varepsilon_{i,j}$ počtu cest mezi oblastmi i a j zjištěné průzkumem od teoretických hodnot získaných gravitačním modelem podle vztahu:

$$\varepsilon_{i,j} = C_{i,j}^{skut} - k \cdot \frac{C_i \cdot A_j}{D_{i,j}^{\alpha}}.$$

Gravitační modely

3) Nyní musíme najít takové hodnoty parametrů modelu, které budou splňovat podmínku:

$$\sum_i \sum_j \varepsilon_{i,j}^2 \rightarrow \min$$

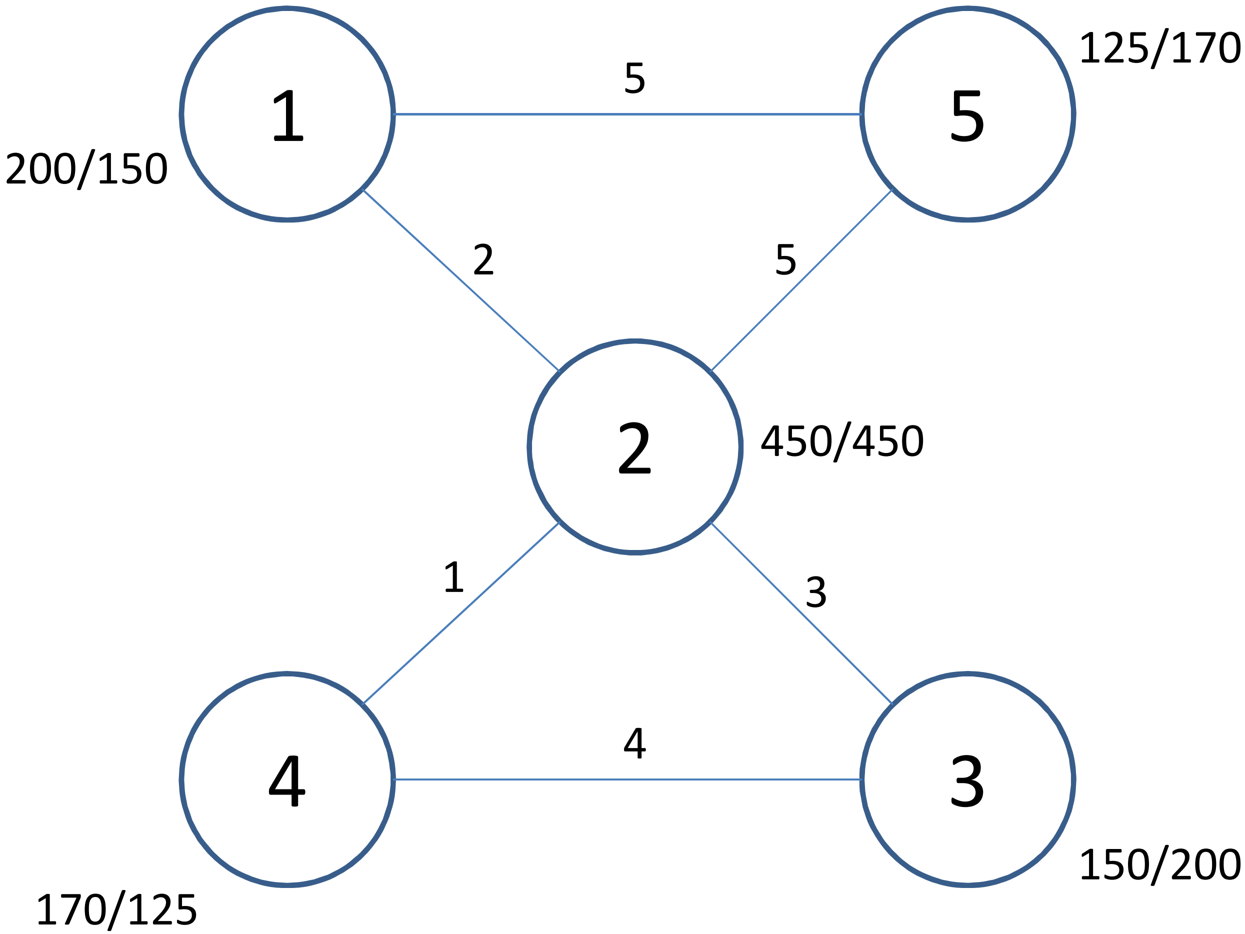
pro všechna *i* a *j* odpovídající relacím, na kterých máme proveden dopravní průzkum.

Gravitační modely

- **Př. 6:** Je dáno 5 oblastí, pro které známe přepravní produktivity a atraktivity. Dále známe vzdálenosti mezi těmito oblastmi.

Oblast	Přepravní produktivita C_i	Přepravní atraktivita A_i
1	200	150
2	450	450
3	150	200
4	170	125
5	125	170
$\sum$	1095	1095

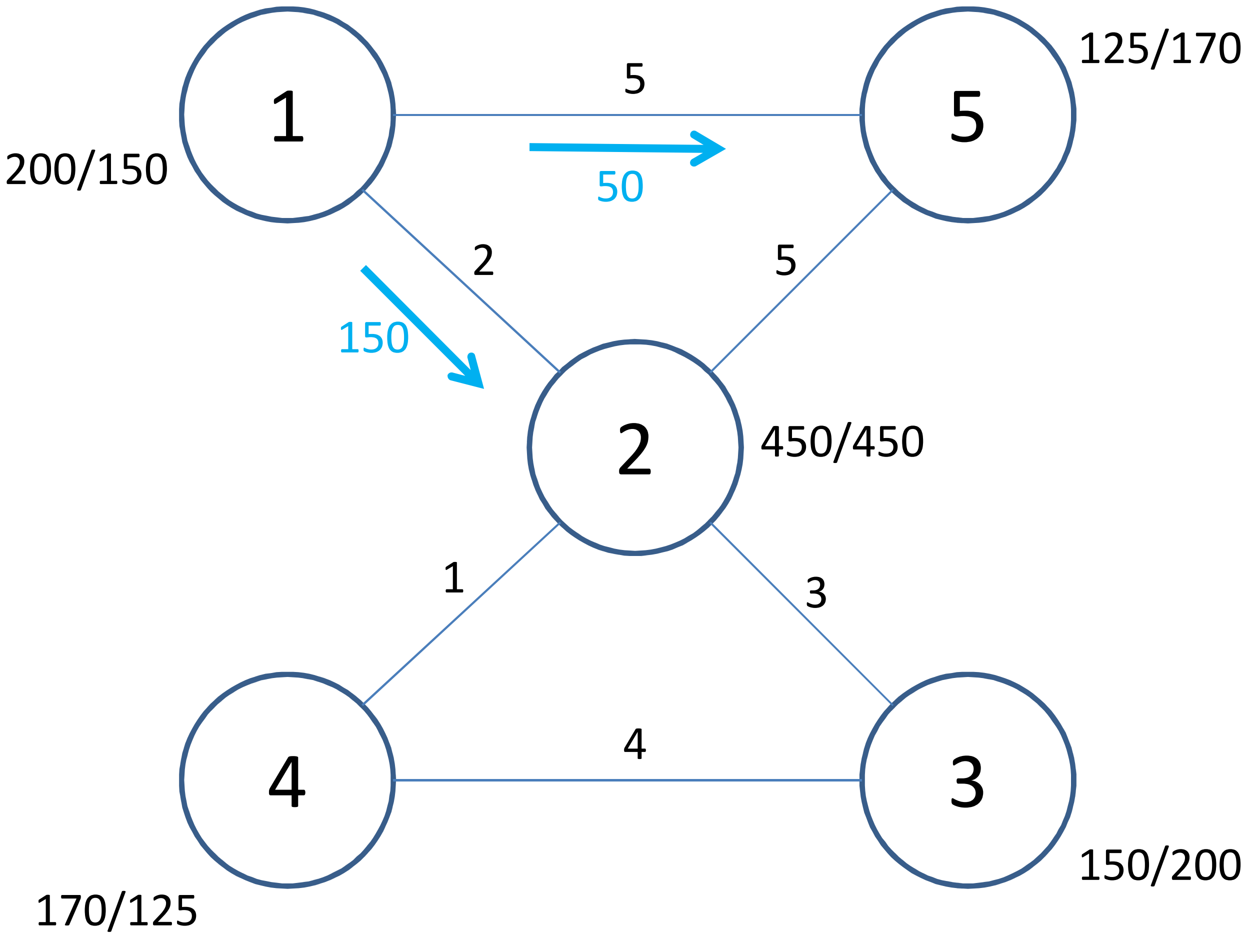
Gravitační modely



Gravitační modely

- Na relacích 1-2 a 1-5 byl proveden dopravní průzkum za účelem zjištění počtu cest mezi těmito oblastmi. Bylo zjištěno, že skutečný počet cest pro relaci 1-2 činí 150 cest/den a relaci 1-5 50 cest/den. Úkolem je odhadnout počty cest pro zbývající relace s využitím gravitačního modelu.

Gravitační modely



Gravitační modely

- V tabulce dole je uvedena celková chyba pro počáteční hodnoty parametrů modelu.

Relace	Přepravní produktivita C_i	Přepravní atraktivita A_j	Vzdálenost $D_{i,j}$	Skutečný tok	Teoretický tok
1-2	200	450	2	150	45000
1-5	200	170	5	50	6800

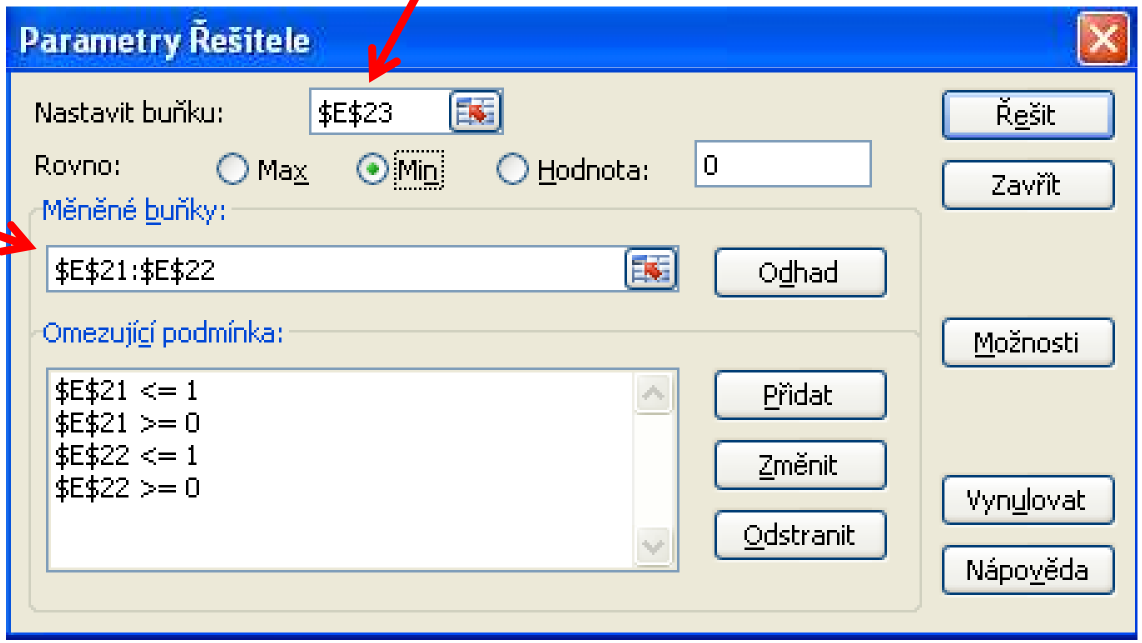
k	**1**
α	**1**
Chyba	2,057E+09

Gravitační modely

- Nyní použijeme doplněk Řešitel.

$$\sum_i \sum_j \varepsilon_{i,j}^2$$

k a *α*



Gravitační modely

Relace	Přepravní produktivita C_i	Přepravní atraktivita A_j	Vzdálenost $D_{i,j}$	Skutečný tok	Teoretický tok
1-2	200	450	2	150	149,9978152
1-5	200	170	5	50	50,00655355

k	**0,001831954**
α	**0,136438698**
Chyba	4,772E-05

$$C_{i,j} \doteq 0{,}00183 \cdot \frac{C_i \cdot A_j}{D_{i,j}^{0{,}13644}}.$$

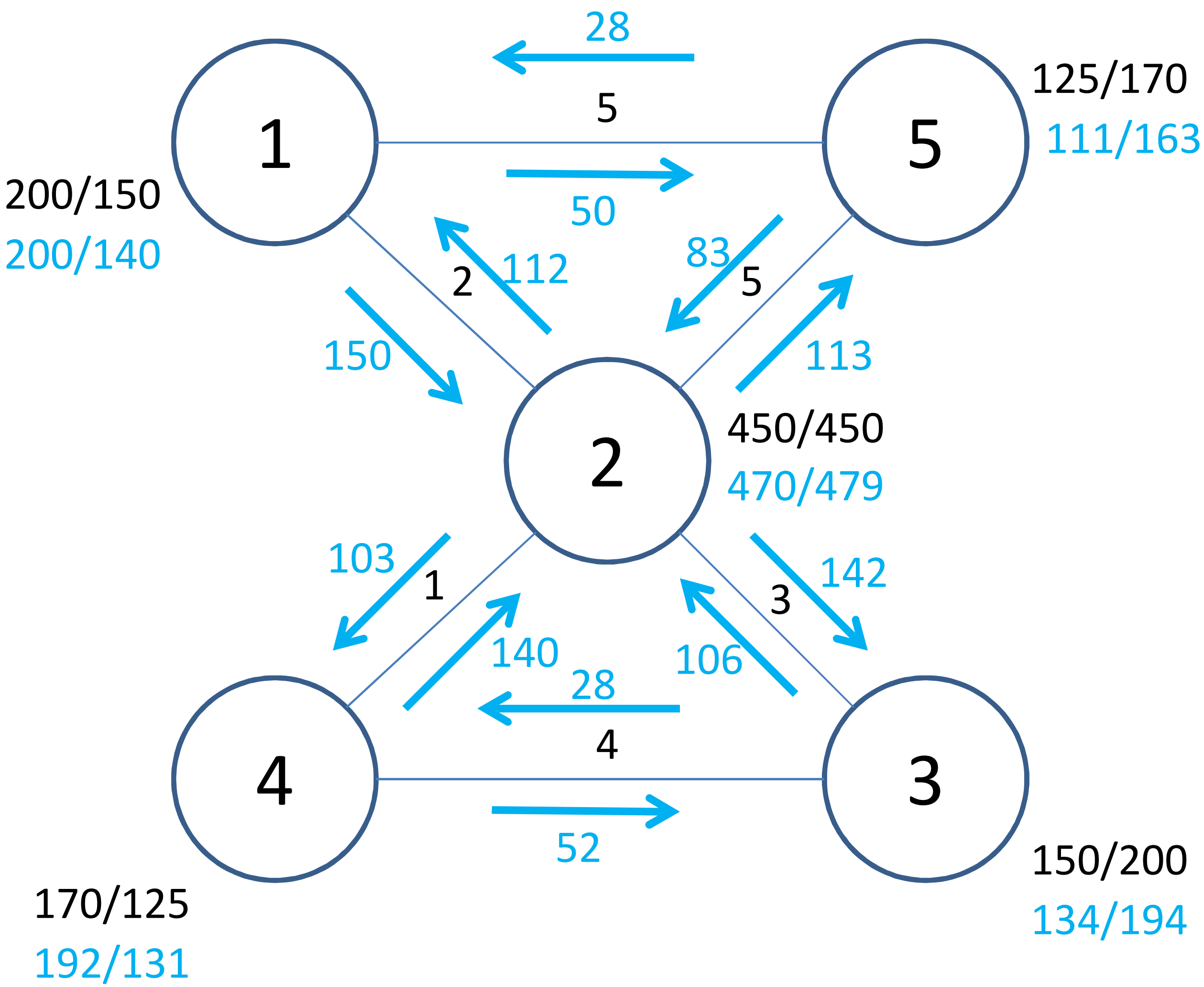
Gravitační modely

- Získali jsme hodnoty parametrů modelu, nyní můžeme stanovit zbývající počty cest pro ostatní relace.

Gravitační modely

Relace	Přepravní produktivita C_i	Přepravní atraktivita A_j	Vzdálenost $D_{i,j}$	Skutečný tok	Teoretický tok
1-2	200	450	2	150	150
1-5	200	170	5	50	50
2-1	450	150	2	-	112
5-1	125	150	5	-	28
2-5	450	170	5	-	113
5-2	125	450	5	-	83
2-4	450	125	1	-	103
4-2	170	450	1	-	140
2-3	450	200	3	-	142
3-2	150	450	3	-	106
4-3	170	200	4	-	52
3-4	150	125	4	-	28

Gravitační modely



Gravitační modely

- Z dosažených výsledků vidíme, že odchylky dopravních produktivit a atraktivit získaných aplikací gravitačního modelu poměrně dobře aproximuje hodnoty vstupující do výpočtu (odchylky jsou maximálně rovny 10%).

Dělba přepravní práce

Dělba přepravní práce

- Dělbou přepravní práce rozumíme způsob rozdělování přepravních objemů mezi jednotlivými druhy doprav – IAD, MHD,…
- Tento problém patří do teorie volby.
- Uvažujme, že máme n variant a každá z nich přináší určitý užitek (nebo ztrátu) u_i.
- Užitek potom ovlivňuje pravděpodobnost volby dané varianty.

Dělba přepravní práce

- Čím vyšší užitek u_i, tím vyšší je pravděpodobnost p_i zvolení příslušné alternativy.
- Jednou z možností je model LOGIT, který stanovuje pravděpodobnosti volby varianty *i* dle vztahu:

$$p_i = \frac{e^{a \cdot u_i}}{\sum_{i=1}^{n} e^{a \cdot u_i}},$$

kde *a* je vhodně zvolený parametr.

Přidělování na síť

Přidělování na síť

- Volba trasy závisí na dvou subjektech – dopravci a přepravci.
- Např. v nákladní železniční dopravě o volbě trasy rozhoduje dopravce.
- V MHD je přepravce (cestující) hlavním faktorem (volí si trasu sám), dopravce ovlivňuje volbu trasy linkovou sítí.
- V IAD o trase rozhoduje výhradně přepravce.

Přidělování na síť

- Grafickým výstupem jsou zátěžové diagramy intenzit (pentlogramy) a kartogramy intenzit na křižovatkách.
- Používají se např.:
  1) Metoda nejkratší trasy.
  2) Metoda přidělení na více tras.
  3) Metoda omezené kapacity.

Přidělování na síť

- ad 1) Metoda nejkratší trasy předpokládá, že při výběru trasy je jednoznačně preferována trasa nejkratší, ostatní trasy nejsou vůbec uvažovány.
- ad 2) Metoda přidělení na více tras vychází z předpokladu, že v 30% případů je vybírána i jiná trasa než nejkratší.

Přidělování na síť

- ad 3) Metoda omezené kapacity bere v úvahu i kapacity komunikací. Nejkratší trasa v oblasti dopravního sedla nemusí být nejkratší trasou i v případě dopravní špičky.