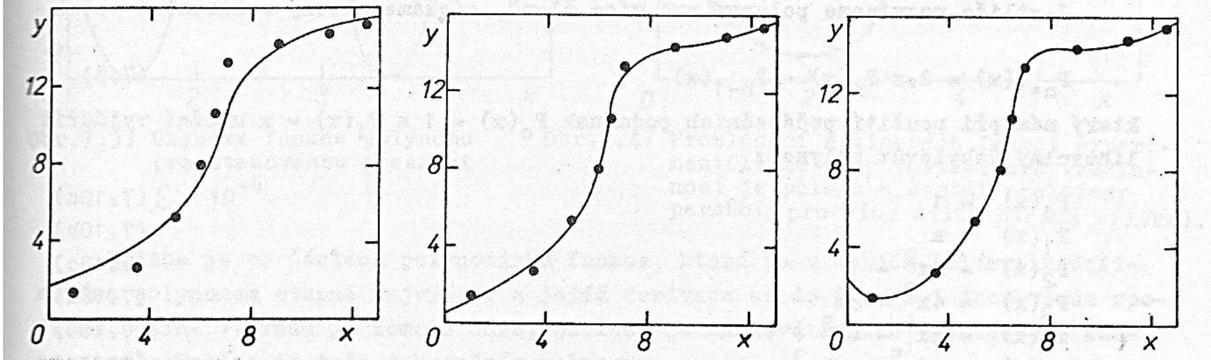
POLYNOMICKÁ REGRESE

Jedná se o regresní model, který je lineární v parametrech, ale popisuje nelineární závislost mezi proměnnými.

$$y = b_0 + b_1x + b_2x^2 + \ldots + b_nx^n$$

kde b_i jsou neznámé parametry, které chceme určit,
n je stupeň polynomu

- Tento model obsahuje pouze jednu nezávisle proměnnou, která se v něm však vyskytuje v různých mocninách, jsou zde vždy všechny mocniny od 1 do n.
- Speciálním případem je přímka – polynom 1. stupně.
- *Tvary polynomických závislostí pro sudý a lichý stupeň polynomu.*



Obr.7.2a: Závislost proložená polynomem třetího stupně.
Obr.7.2b: Závislost proložená polynomem čtvrtého stupně.
Obr.7.2c: Závislost proložená polynomem šestého stupně.

MULTIKOLINEARITA

- Multikolinearita – vysoké hodnoty párových korelačních koeficientů mezi vysvětlujícími proměnnými – přibližná rovnoběžnost sloupcových vektorů v matici **X**.
- Početní problémy během MNČ: špatná podmíněnost matice $\mathbf{X}^T\mathbf{X}$ kvůli hodnotám vlastních čísel λ blízkých nule; nelze provést inverzi matice = regresní model není jednoznačně řešitelný.
- Statistické problémy: neúměrně vysoké rozptyly regresních koeficientů vedoucí k jejich nespolehlivému určení; nestabilita odhadů regresních koeficientů.
- Multikolinearita je porušením předpokladů pro MNČ!

IDENTIFIKACE MULTIKOLINEARITY

Číslo podmíněnosti K:

$$K = \frac{\lambda_{max}}{\lambda_{min}} \quad \text{, kde}$$

λ_{max} a λ_{min} jsou maximální a minimální vlastní čísla.
Hodnota K > 1000 ukazuje na silnou multikolinearitu.

VI-faktor (Variance Inflation Factor):

$$VIF = \frac{1}{1 - R_j^2}$$

Je-li VIF > 10, jde o silnou multikolinearitu.

Scottova testační charakteristika M_T:

$$M_T = \frac{\frac{F_R}{t_S} - 1}{\frac{F_R}{t_S} + 1}, kde$$

$\underline{F_R}$ je testační kritérium ze čtverců testačních statistik t_i (test významnosti regresního koeficientu) a $\underline{t_S}$ je průměrná hodnota čtverců testačních charakteristik t_i.

Posouzení:

a) $M_T > 0{,}8$ – model z hlediska multikolinearity nevyhovuje, je nezbytná úprava modelu
b) $0{,}33 < M_T < 0{,}8$ – model málo vyhovující, úprava však není nezbytná,
c) $M_T < 0{,}33$ – model není multikolinearitou ovlivněn.

POSTUP ŘEŠENÍ polynomické regrese:

- určení vhodného stupně polynomu.
- redukce nadbytečných členů polynomu.

URČENÍ STUPNĚ POLYNOMU

Pro jednotlivé stupně polynomu sledujeme charakteristiky MEP, AIC a R_p^2. Nejmenším hodnotám MEP a AIC a největší R_p^2 odpovídá nejlepší stupeň polynomu. Doporučuje se používat spíše nižších stupňů polynomu, polynomy vyššího stupně mají sklon k numerické nestabilitě, která se projevuje silným "rozkmitáním" křivky a špatnou predikční schopností.

<u>Redukce členů polynomu:</u>
<u>Metoda racionálních hodností (korekce hodností)</u>

- Vede k vychýleným odhadům parametrů b_i s nižším rozptylem a pokud možno minimálním vychýlením, které jsou méně citlivé na špatnou podmíněnost matice $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ nebo **R**.
- Postup výpočtu je stejný jako u MNČ (je hledán minimální RSC), vkládá se však omezení na velikost vlastních čísel λ!
- <u>Parametr omezení na vlastní čísla P:</u> P = 0 – MNČ; hodnoty větší než nula potlačí komponenty vzniklé rozkladem na vlastní čísla odpovídající nejmenším vlastním číslům. Doporučuje se hodnota nejvýše kolem 0,1.
- Po vynechání <u>*k*</u> malých vlastních čísel λ z korelační matice **R** se vypočtou nové odhady parametrů b_i s menším rozptylem, které jsou však vychýlené. Tyto vychýlené odhady však vyhovují lépe než nevychýlené odhady získané pomocí MNČ.

VÍCEROZMĚRNÁ LINEÁRNÍ REGRESE

- Pokud měřené hodnoty *y* závisejí na více faktorech *x*, není vždy možné provést takovou sérii pokusů, při nichž by se měnila pouze hodnota jediného faktoru a ostatní *x* zůstaly konstantní. To by umožňovalo sledovat vliv jediného faktoru jednoduchou regresí.
- Experimentální výsledky lze zpracovávat ve vztahu k současné změně více faktorů pomocí **vícerozměrné (vícenásobné) regrese**.

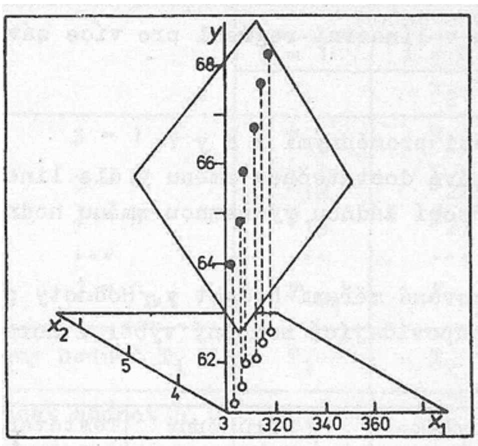
Jsou-li hodnoty *y* lineárně závislé na několika nezávislých proměnných x_i, tj. platí-li

$$\mathbf{y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k}$$

b_i … regresní parametry o počtu k; (k = 1 – jednoduchá lineární regrese)

můžeme pro výpočet odhadů koeficientů b_i použít metodu nejmenších čtverců (jsou-li splněny předpoklady pro její použití).

- Absolutní člen b_0 je průsečíkem regresní nadroviny s osou y.
- Parametry b_i jsou směrnice regresní nadroviny ve směru x_i a jsou nazývány parciálními regresními parametry (koeficienty).



Při vícerozměrné regresi se používá stejný postup jako u jednorozměrné LR, při regresní diagnostice se používají k posouzení kvality regresního modelu také:

- parciální grafy,
- vícenásobný korelační koeficient,
- parciální korelační koeficienty.