Regressions- och tidsserieanalys

Föreläsning 7 - Icke-linjär regression. Polynom- och exponentiella samband

Mattias Villani

Statistiska institutionen
Stockholms universitet

Institutionen för datavetenskap
Linköpings universitet

mattiasvillani.com @matvil mattiasvillani

Översikt

- Polynomregression
- Exponentiella modeller

Kvadratisk regression

- **Kvadratisk regression**

$$y = a + b_1 x + b_2 x^2$$

- Samma som **multipel regression med två förklarande variabler**:
  - $x_1 = x$
  - $x_2 = x^2$
- **Populationsmodell**:

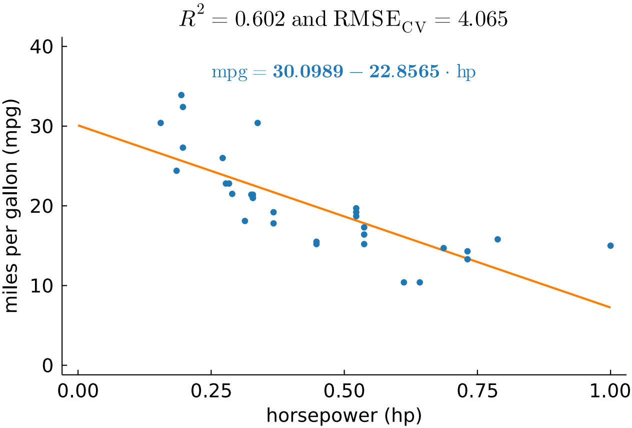
$$y = \alpha + \beta_1 x + \beta_2 x^2 + \varepsilon$$

- **Minsta-kvadratmetoden** för att beräkna a, b_1 och b_2!
- Kvadratisk regression **icke-linjär i x**, men linjär i α, β_1 och β_2.

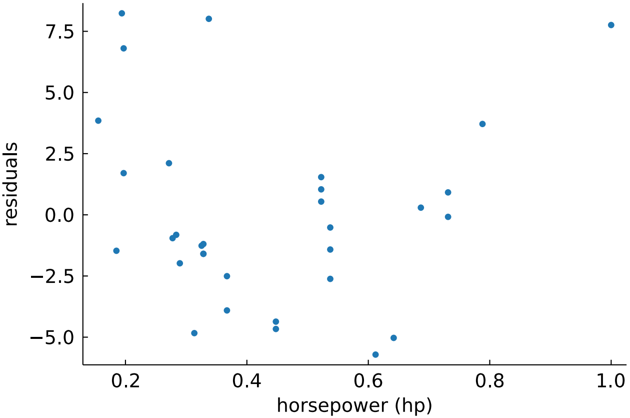
Kvadratisk regression - excel

	A	B	C	D
1		mpg (y)	hp (x)	x^2
2	Mazda RX4	21.000	0.328	0.108
3	Mazda RX4 Wag	21.000	0.328	0.108
4	Datsun 710	22.800	0.278	0.077
5	Hornet 4 Drive	21.400	0.328	0.108
6	Hornet Sportabo	18.700	0.522	0.273
7	Valiant	18.100	0.313	0.098
8	Duster 360	14.300	0.731	0.535
9	Merc 240D	24.400	0.185	0.034
10	Merc 230	22.800	0.284	0.080
11	Merc 280	19.200	0.367	0.135
12	Merc 280C	17.800	0.367	0.135
13	Merc 450SE	16.400	0.537	0.289
14	Merc 450SL	17.300	0.537	0.289
15	Merc 450SLC	15.200	0.537	0.289
16	Cadillac Fleetwo	10.400	0.612	0.374
17	Lincoln Continen	10.400	0.642	0.412
18	Chrysler Imperia	14.700	0.687	0.471
19	Fiat 128	32.400	0.197	0.039
20	Honda Civic	30.400	0.155	0.024

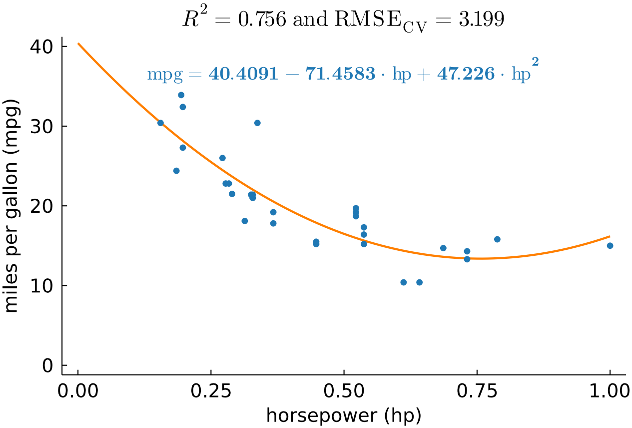
Cars data - linjär regression mot hp



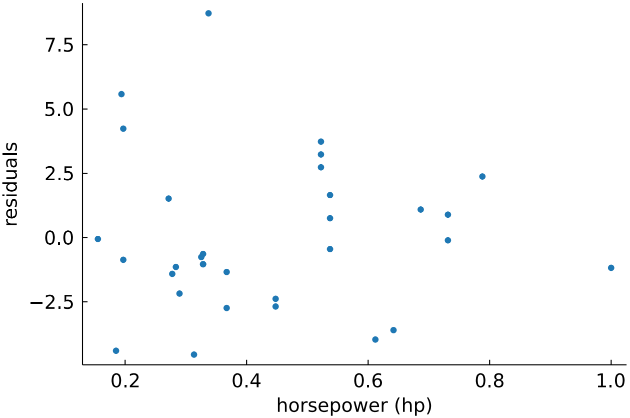
Cars data - residualer linjär regression



Cars data - kvadratisk regression mot hp



Cars data - residualer kvadratisk regression



Tolkningar av parametrar i kvadratisk regression

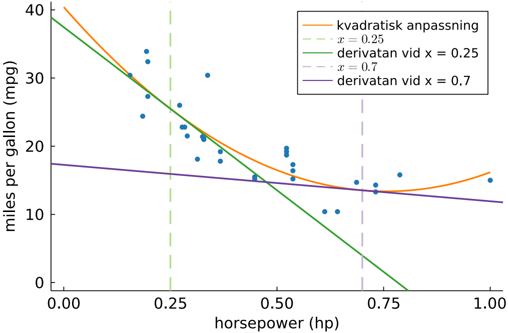
- **Kvadratisk regression**

$$y = a + b_1 x + b_2 x^2$$

- **Regressionskoefficienterna** tolkas som **derivator**:

$$\frac{dy}{dx} = b_1 + 2b_2 \cdot x$$

- Effekten av en liten förändring Δx i x beror på x själv:



Polynomregression

- **Polynomregression**

$$y = a + b_1 x + b_2 x^2 + \ldots + b_k x^k$$

- Polynomregression av ordning $k$är detsamma som multipel regression med k förklarande variabler:
  - $x_1 = x$
  - $x_2 = x^2$
  - $\vdots$
  - $x_k = x^k$
- **Populationsmodell**:

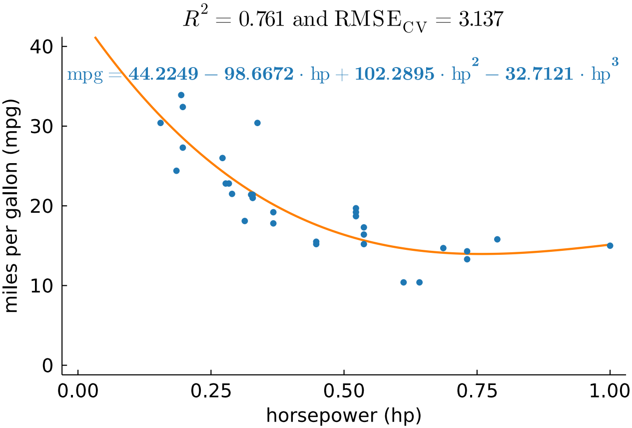
$$y = \alpha + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k + \varepsilon$$

- **Minsta-kvadratmetoden** kan användas för att beräkna $a, b_1, \ldots b_k$!
- Polynomregression är **icke-linjär i x**, men linjär i α, $\beta_1, \ldots \beta_k$.

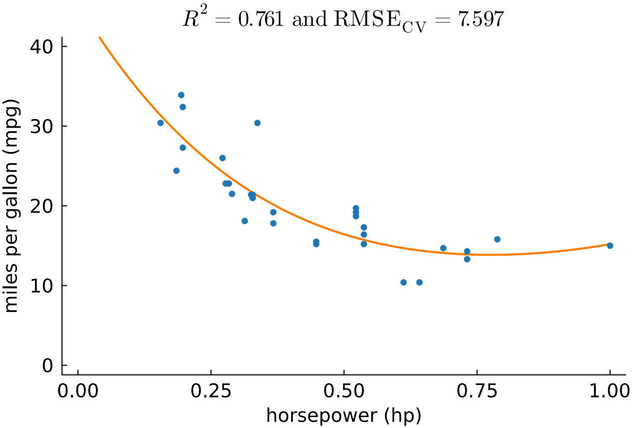
Polynomregression - excel

	A	B	C	D	E	F
1		mpg (y)	hp (x)	x^2	x^3	x^4
2	Mazda RX4	21.000	0.328	0.108	0.035	0.012
3	Mazda RX4 Wag	21.000	0.328	0.108	0.035	0.012
4	Datsun 710	22.800	0.278	0.077	0.021	0.006
5	Hornet 4 Drive	21.400	0.328	0.108	0.035	0.012
6	Hornet Sportabo	18.700	0.522	0.273	0.143	0.074
7	Valiant	18.100	0.313	0.098	0.031	0.010
8	Duster 360	14.300	0.731	0.535	0.391	0.286
9	Merc 240D	24.400	0.185	0.034	0.006	0.001
10	Merc 230	22.800	0.284	0.080	0.023	0.006
11	Merc 280	19.200	0.367	0.135	0.049	0.018
12	Merc 280C	17.800	0.367	0.135	0.049	0.018
13	Merc 450SE	16.400	0.537	0.289	0.155	0.083
14	Merc 450SL	17.300	0.537	0.289	0.155	0.083
15	Merc 450SLC	15.200	0.537	0.289	0.155	0.083
16	Cadillac Fleetwo	10.400	0.612	0.374	0.229	0.140
17	Lincoln Continer	10.400	0.642	0.412	0.264	0.170
18	Chrysler Imperia	14.700	0.687	0.471	0.324	0.222
19	Fiat 128	32.400	0.197	0.039	0.008	0.002

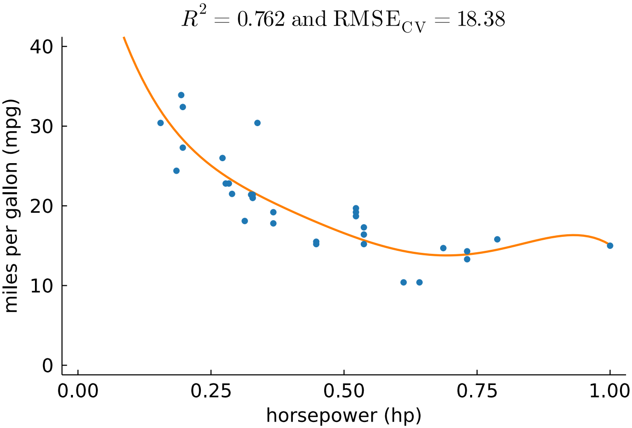
Cars data - kubisk regression mot hp



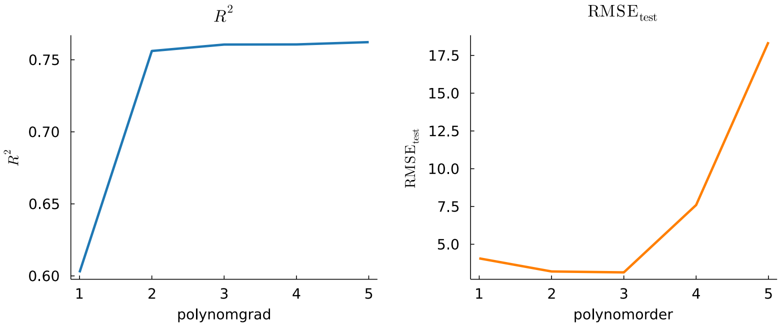
Cars data - polynomregression ordning 4



Cars data - polynomregression ordning 5



Cars data - R^2 och RMSE-CV($K = 4$) på testdata



Exponentiella samband

- Du sätter in 200 kr på banken till 5% årsränta. Utveckling:

$$1 \text{ år: } 200 \cdot 1.05 = 210.000 \text{ kr}$$

$$2 \text{ år: } 200 \cdot 1.05^2 = 220.500 \text{ kr}$$

$$3 \text{ år: } 200 \cdot 1.05^3 = 231.525 \text{ kr}$$

- Efter x år: $200 \cdot 1.05^x$. **Exponentiell tillväxt**. Samma procentuella ökning varje år.
- **Exponentiellt samband**

$$y = a \cdot b^x$$

- a är det **initiala** beloppet eller storheten.
- b bestämmer **tillväxttakten**

$$b > 1 \text{ ökande}$$

$$b < 1 \text{ minskade}$$

$$b = 1 \text{ konstant}$$

Exponentiell regression

- **Exponentiell regression**:

$$y = a \cdot b^x$$

- Logaritmera (10-logaritmer) båda sidor

$$\underbrace{\log y}_{\tilde{y}} = \underbrace{\log a}_{\tilde{a}} + \underbrace{\log b}_{\tilde{b}} \cdot x$$

$$\tilde{y} = \tilde{a} + \tilde{b}x$$

$$\tilde{a} = \log a$$

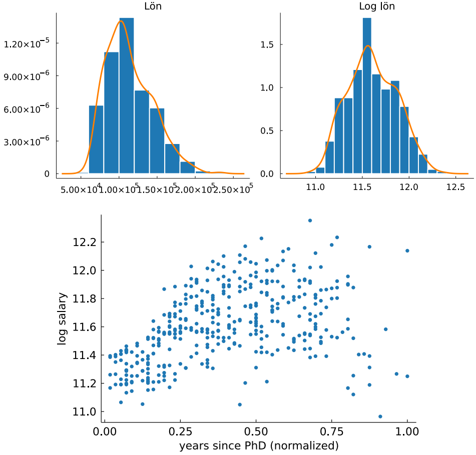
$$\tilde{b} = \log b$$

- Skatta $\tilde{a}$ och $\tilde{b}$ med **minsta-kvadrat** med $\tilde{y} = \log y$!
- Skattningar för a och b fås genom

$$a = 10^{\tilde{a}} \qquad \text{och} \qquad b = 10^{\tilde{b}}$$

Exponentiell regression

- **Responsvariabler** med **enbart positiva värden** (t ex lön):
  - **Normalfördelning** ofta **opassande** pga **skevhet**.
  - **kan ge prediktioner för y som är negativa**.



Exponentiell regression

- **Populationsmodell**:
$$y = \alpha \cdot \beta^x \varepsilon$$
- **Logaritmen av feltermen** ε är **normalfördelad.**
- Vi säger att feltermen ε är **lognormal** fördelad. Innebär $\varepsilon > 0$.
- Logaritmera!
$$\tilde{y} = \tilde{\alpha} + \tilde{\beta} \cdot x + \tilde{\varepsilon}, \qquad \tilde{\varepsilon} \sim N\left(0, \sigma^2_{\tilde{\varepsilon}}\right).$$
- Notera att $\tilde{\beta} = 0$ innebär att $\beta = 1$.
- Så t-test för $H_0 : \tilde{\beta} = 0$ är **test för konstant tillväxt**.
- **Prediktion** för $x = x_0$:
$$\hat{y} = a \cdot b^{x_0} = 10^{\tilde{a}} \cdot (10^{\tilde{b}})^{x_0} = 10^{\tilde{a} + \tilde{b} x_0}$$