Metoda nejmenších čtverců – lineární model

- sada naměřených hodnot $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ – nezávislé proměnné

  $\boldsymbol{y} = (y_1, y_2, ..., y_n)$ – závislé proměnné $y_i \in N(\mu_i, \sigma_i)$

- modelová funkce $\boldsymbol{\lambda} = \boldsymbol{A\theta}$ – modelujeme závislost $y(x)$

$$\lambda(x_i, \boldsymbol{\theta}) = \sum_{j=1}^{m} a_j(x_i)\theta_j = \sum_{j=1}^{m} A_{ij}\theta_j$$

  $\boldsymbol{\theta} = \theta_1, \theta_2, ...\theta_m$ – parametry modelové závislosti

- minimalizujeme tzv. „chí kvadrát“

$$\chi^2(\boldsymbol{\theta}|\boldsymbol{y}) = \sum_{i=1}^{n} \frac{\left(y_i - \sum_{j=1}^{m} A_{ij}\theta_j\right)^2}{\sigma_i^2}$$

pokud $\sigma_i = \sigma$ potom

$$\boxed{\chi^2 = \frac{1}{\sigma^2}(\boldsymbol{y} - \boldsymbol{A\theta})^T(\boldsymbol{y} - \boldsymbol{A\theta})}$$

obecně $V_{ij} = \mathrm{cov}(y_i, y_j)$

$$\boxed{\chi^2 = (\boldsymbol{y} - \boldsymbol{A\theta})^T \boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{A\theta})}$$

Metoda nejmenších čtverců – lineární model

- minimalizujeme tzv. „chí kvadrát“ $$\chi^2(\boldsymbol{\theta}|\boldsymbol{y}) = \sum_{i=1}^{n} \frac{\left(y_i - \sum_{j=1}^{m} A_{ij}\theta_j\right)^2}{\sigma_i^2}$$

$$\frac{\partial\chi^2}{\partial\theta_k} = 0$$

$$\frac{\partial\chi^2}{\partial\theta_k} = -\frac{2}{\sigma^2}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{m} A_{ij}\theta_j\right)A_{ik} \qquad \Rightarrow \sum_{i=1}^{n} A_{ik}y_i = \sum_{i=1}^{n}\sum_{j=1}^{m} A_{ik}A_{ij}\theta_j$$

$$\boldsymbol{A}^T\boldsymbol{y} = \boldsymbol{A}^T\boldsymbol{A}\boldsymbol{\theta}$$

- odhad parametrů $\hat{\theta}$

pro $\sigma_i = \sigma$ $$\hat{\boldsymbol{\theta}} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{y} \equiv \boldsymbol{B}\boldsymbol{y}$$

obecně $$\hat{\boldsymbol{\theta}} = (\boldsymbol{A}^T\boldsymbol{V}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{V}^{-1}\boldsymbol{y} \equiv \boldsymbol{B}\boldsymbol{y}$$

- odhad kovariance parametrů $\hat{\theta}$

$$U_{ij} = \mathrm{cov}(\theta_i, \theta_j)$$

$$\boldsymbol{U} = \boldsymbol{B}\boldsymbol{V}\boldsymbol{B}^T$$

$$U_{ij}^{-1} = \frac{1}{2}\left[\frac{\partial^2\chi^2}{\partial\theta_i\partial\theta_j}\right]_{\theta=\hat{\theta}}$$

Metoda nejmenších čtverců – fit polynomu

- obecný polynom $\lambda(x_i, \boldsymbol{\theta}) = \sum_{j=0}^{m} A_{ij}\theta_j = \sum_{j=0}^{m} \theta_j x_i^j$

$$\boldsymbol{\theta} = \theta_1, \theta_2, \ldots \theta_m$$

$$\sigma_i = \sigma$$

$$A_{ij} = x_i^j$$

$$\boldsymbol{A} = \begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^m \\ 1 & x_2 & x_2^2 & \ldots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^m \end{pmatrix} \qquad n \times m$$

$$\boldsymbol{A}^T\boldsymbol{A} = \begin{pmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \ldots & \sum_{i=1}^n x_i^m \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \ldots & \sum_{i=1}^n x_i^{m+1} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum_{i=1}^n x_i^m & \sum_{i=1}^n x_i^{m+1} & \sum_{i=1}^n x_i^{m+2} & \ldots & \sum_{i=1}^n x_i^{m+m} \end{pmatrix} \qquad m \times m$$

$$\boxed{\begin{aligned} \hat{\boldsymbol{\theta}} &= (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{y} \\ \boldsymbol{U} &= \sigma^2(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\left[(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\right]^T \end{aligned}}$$

– operace s maticemi (Matlab)

Metoda nejmenších čtverců – fit paraboly

- polynom 2. stupně $\lambda(x_i, \boldsymbol{\theta}) = \sum_{j=0}^{m} A_{ij}\theta_j = \theta_0 + \theta_1 x_i + \theta_2 x_i^2$

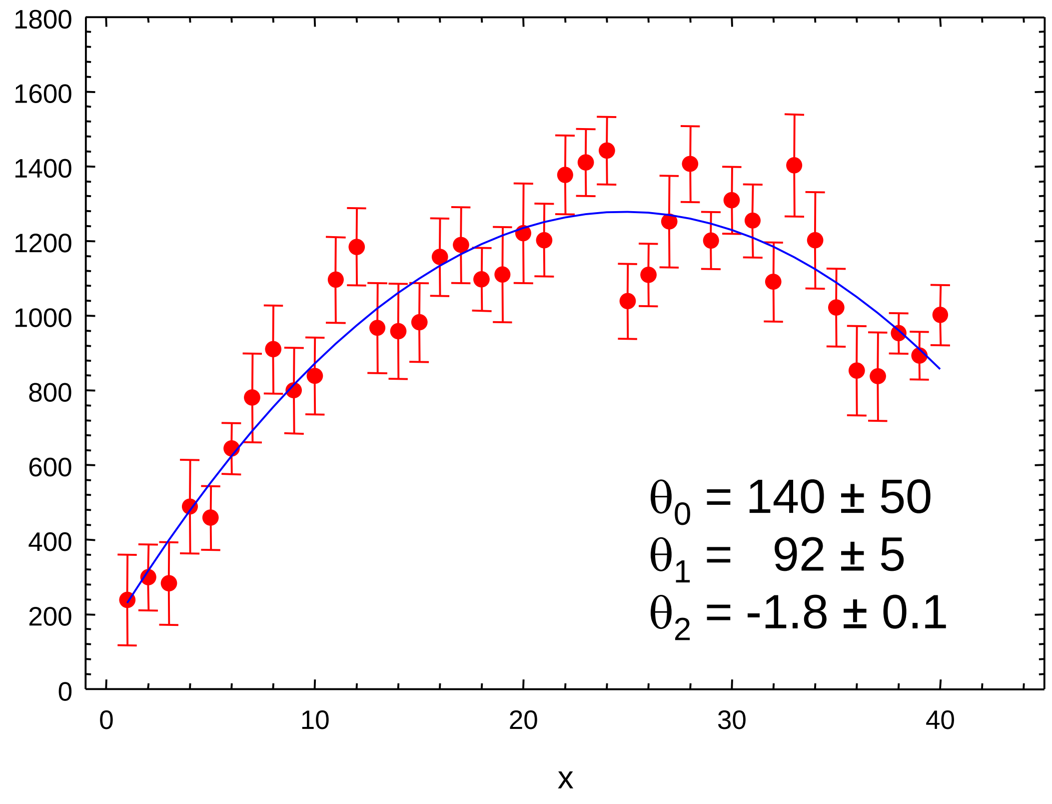
$$\boldsymbol{A} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix}$$

$$\boldsymbol{A}^T \boldsymbol{V}^{-1} \boldsymbol{A} = \begin{pmatrix} \sum_{i=1}^n \frac{1}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^3}{\sigma_i^2} \\ \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^3}{\sigma_i^2} & \sum_{i=1}^n \frac{x_i^4}{\sigma_i^2} \end{pmatrix}$$

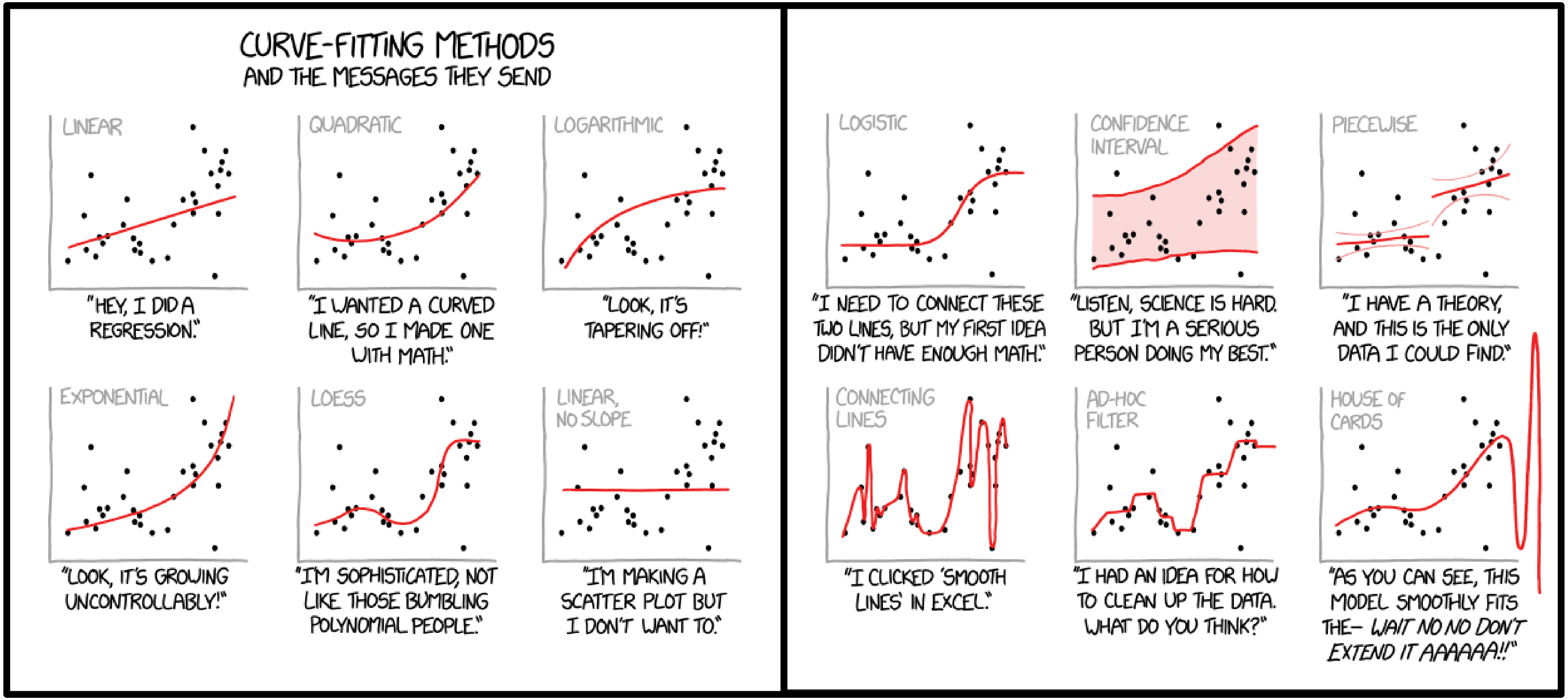
$$\hat{\boldsymbol{\theta}} = (\boldsymbol{A}^T \boldsymbol{V}^{-1} \boldsymbol{A})^{-1} \boldsymbol{A}^T \boldsymbol{V}^{-1} \boldsymbol{y} \equiv \boldsymbol{B}\boldsymbol{y}$$

$$\boldsymbol{U} = \boldsymbol{B}\boldsymbol{V}\boldsymbol{B}^T$$

$$U_{ij} = \mathrm{cov}(\theta_i, \theta_j)$$



χ^2 test kvality fitu



xkcd.com/2048/

χ^2 test kvality fitu

- sada naměřených hodnot $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ – nezávislé proměnné

  $\boldsymbol{y} = (y_1, y_2, ..., y_n)$ – závislé proměnné $y_i \in N(\mu_i, \sigma_i)$

- modelová funkce $\lambda(x, \boldsymbol{\theta})$ – modelujeme závislost $y(x)$

  $\boldsymbol{\theta} = \theta_1, \theta_2, ...\theta_m$ – parametry modelové závislosti

- testovací statistika

$$\chi^2(\boldsymbol{\theta}|\boldsymbol{y}) = \sum_{i=1}^{n} \frac{(y_i - \lambda(x_i|\boldsymbol{\theta}))^2}{\sigma_i^2}$$

náhodná proměnná χ^2 s rozdělením
o *n* - *m* stupních volnosti

$$\chi^2 \in \chi^2(n - m)$$

χ^2 rozdělení

- hustota pravděpodobnosti $$\boxed{f(y|n) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} \qquad y \in [0,\infty)\,, \qquad n = 1, 2, \ldots}$$

počet stupňů volnosti n

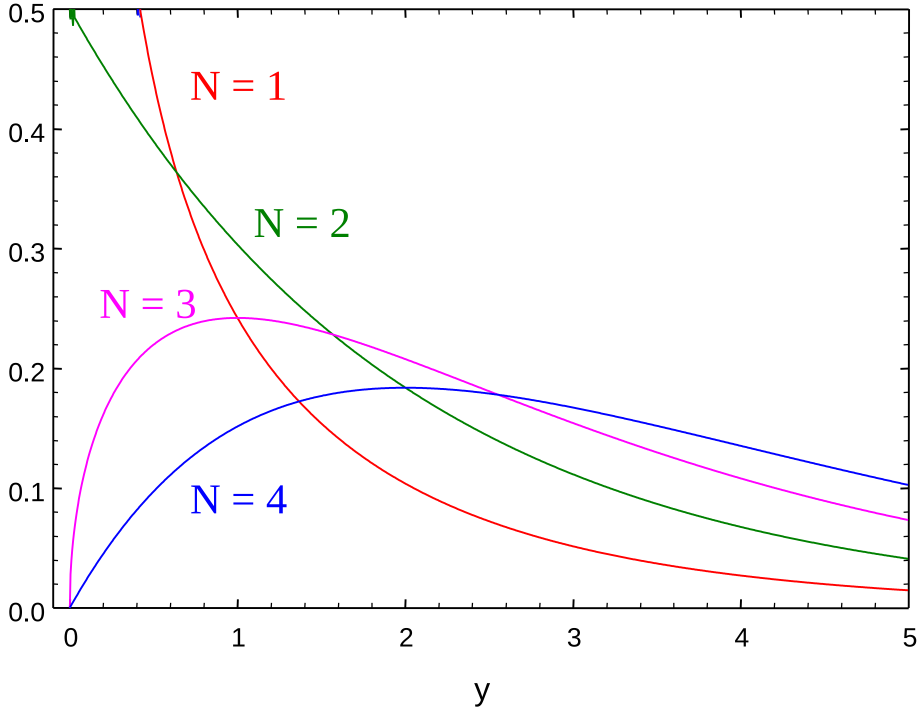
gama funkce $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} \mathrm{d}t$

$$z_i \in N(0,1)$$

$$y = \sum_{i=1}^{n} z_i \quad \Rightarrow \quad y \in f(n) \text{ resp. } \chi^2(n)$$

$$x_i \in N(\mu_i, \sigma_i)$$

$$y = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2} \quad \Rightarrow \quad y \in f(n) \text{ resp. } \chi^2(n)$$



- momenty χ^2 rozdělení

$$E[y] = n$$

$$V[y] = 2n$$

χ^2 test kvality fitu

- χ^2 rozdělení $f(y|n-m) = \frac{1}{2^{\frac{n-m}{2}}\Gamma(\frac{n-m}{2})} y^{\frac{n-m}{2}-1} e^{-\frac{y}{2}}$

$$\boxed{P\left[y \geq \chi_0^2\right] = \int_{\chi_0^2}^{\infty} f(y|n-m)\mathrm{d}y}$$

- počet stupňů volnosti $n-m$

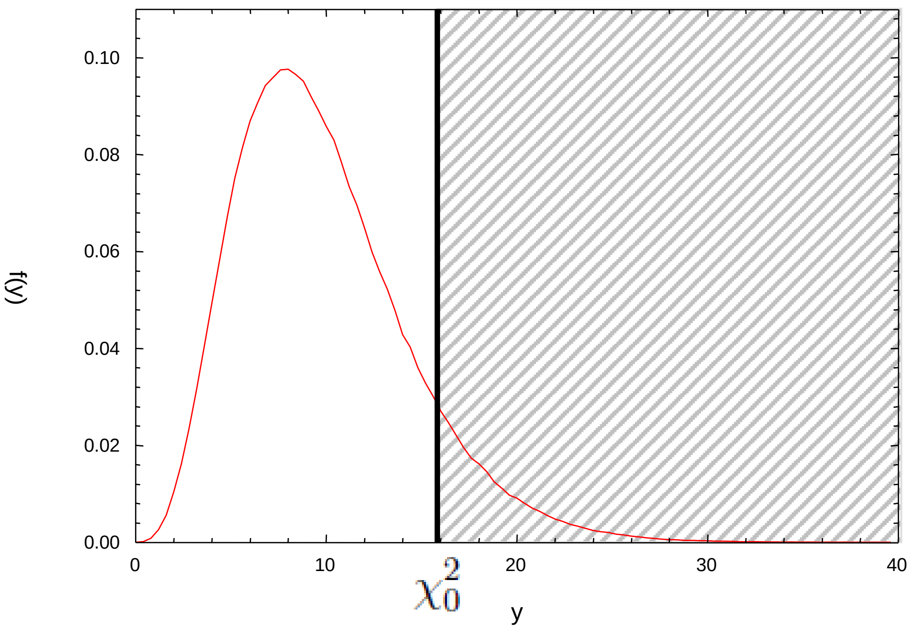
$$E[\chi^2] = n-m$$

$$V[\chi^2] = 2(n-m)$$

- χ^2 na počet stupňů volnosti $\chi^2/(n-m)$

$$E\left[\chi^2/(n-m)\right] = 1$$

$$V\left[\chi^2/(n-m)\right] = 2/(n-m)$$



- nulová hypotéza — naměřené hodnoty $y_i \in N(\mu_i, \sigma_i)$ navzájem nezávislé
  modelová funkce $\lambda(x, \boldsymbol{\theta})$ správně vystihuje závislost $y(x)$

- pokud je P-hodnota $P\left[y \geq \chi_0^2\right] < \alpha$, zamítneme nulovou hypotézu

- α hladina signifikance (typicky 0.05 nebo 0.01)

χ^2 test kvality fitu

- tabulka hodnot $P\left[y \geq \chi^2\right]$ pro počet stupňů volnosti $k = 1 - 10$

Počet stupňů volnosti	χ^2										
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.87	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
P[$y \geq \chi^2$]	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001

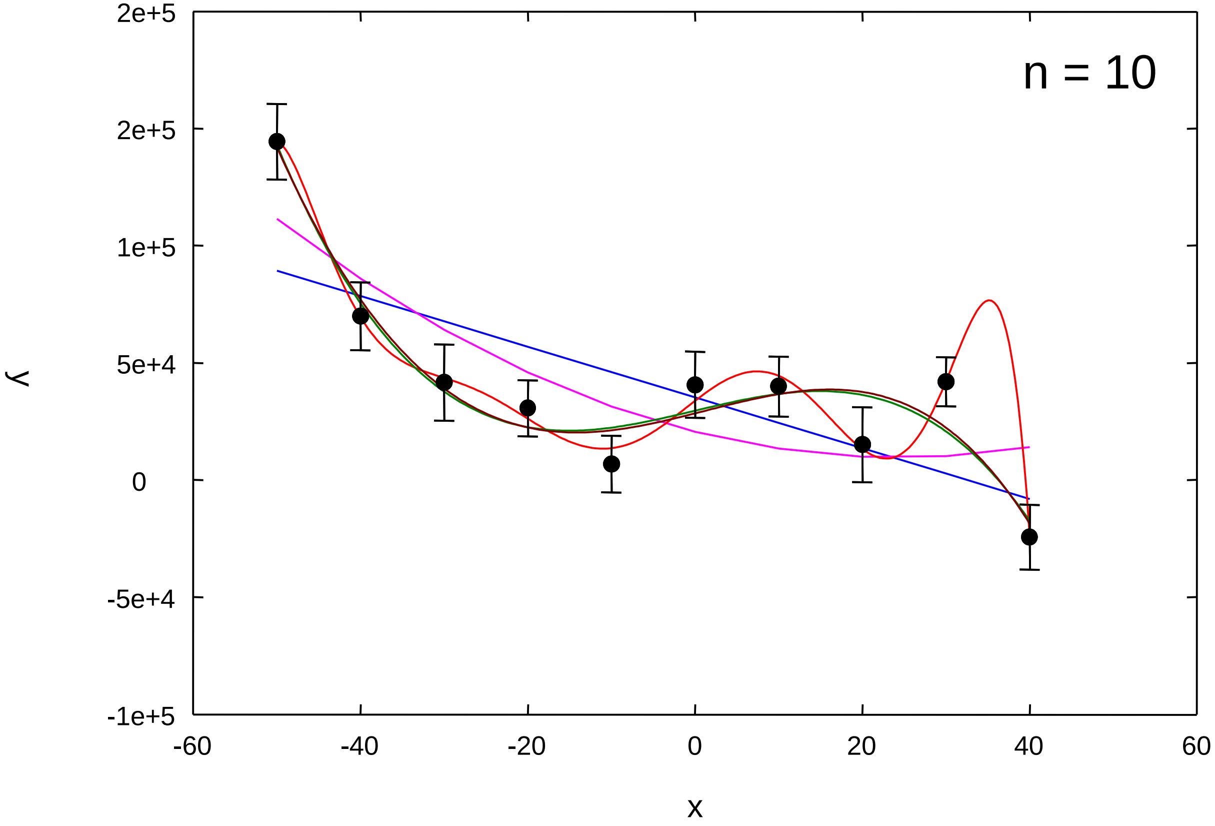
- pro počet stupňů volnosti $k > 50$ rozdělení $\chi^2(k)$ konverguje k normálnímu rozdělení $N(k, \sqrt{2k})$

χ^2 test kvality fitu – polynom

***m* = 2**, χ^2 = 47.04
$\chi^2 / (n - m)$ = 5.88
P = 5.0 × 10^{-6}

***m* = 3**, χ^2 = 36.47
$\chi^2 / (n - m)$ = 5.21
P = 1.4 × 10^{-6}

***m* = 4**, χ^2 = 9.06
$\chi^2 / (n - m)$ = 1.51
P = 0.68



***m* = 5**, χ^2 = 8.60
$\chi^2 / (n - m)$ = 1.72
P = 0.64

***m* = 9**, χ^2 = 7.56
$\chi^2 / (n - m)$ = 0.84
P = 0.53

χ^2 test kvality fitu – binovaná data

$$\chi^2 = \frac{\sum_{i=1}^{n} (y_i - f(\theta_i))^2}{f(\theta_i)}$$

$$n = 201$$

$$\nu = 201 - 5 = 196$$

$\chi^2(\nu)$ $\quad \mu = \nu = 196$

$$\sigma = \sqrt{2\nu} = 20$$

$\chi^2(\nu)/\nu$ $\quad E\left[\chi^2(\nu)/\nu\right] = 1$

$$\sigma_{\chi^2(\nu)/\nu} = 0.1$$