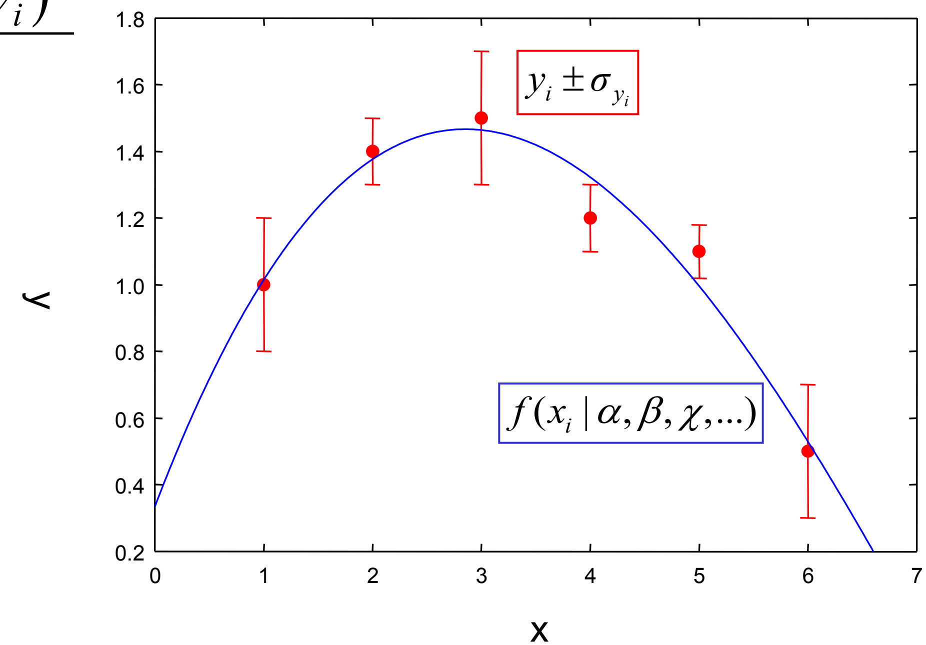
Metoda nejmenších čtverců

- Metoda početní interpolace.
- Používá se pro získání odhadů parametrů $(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}, \ldots)$:

1) Zkonstruujeme veličinu

$$\chi^2(\alpha, \beta, \gamma, \ldots) = \sum_{i=1}^{N} \frac{\left(f(x_i \mid \alpha, \beta, \gamma, \ldots) - y_i\right)^2}{\sigma_{y_i}^2}$$

2) Hledáme minimum $\chi^2(\alpha,\beta,\gamma,\ldots)$.



Metoda nejmenších čtverců: chyby v x a y

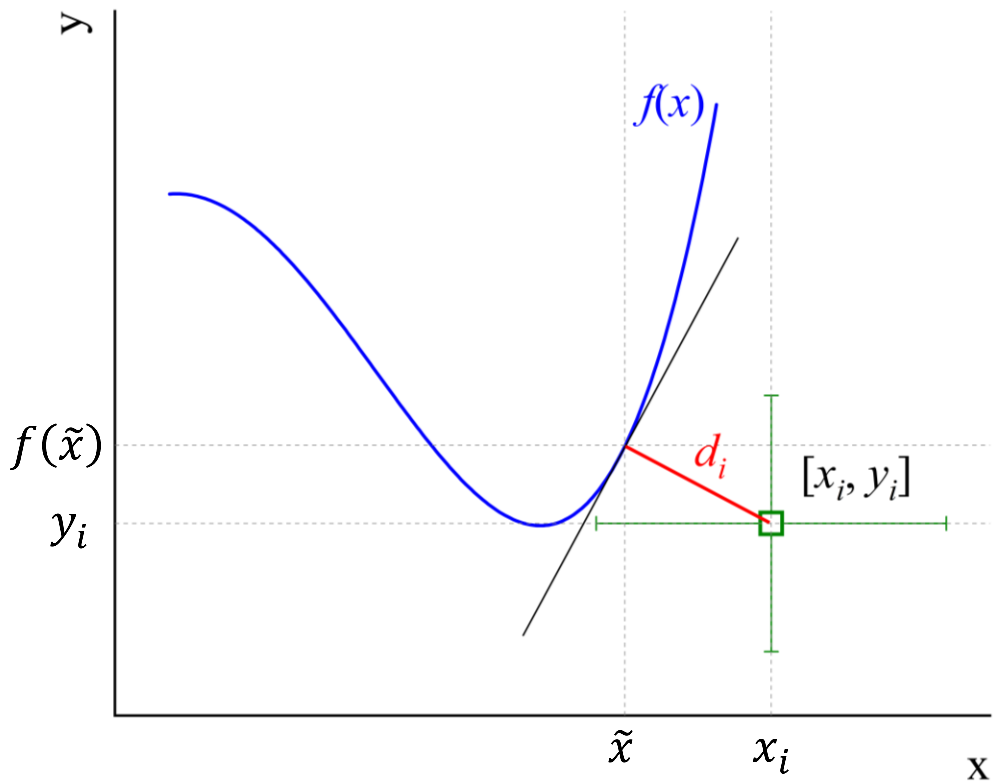
- Uvažujeme nyní chyby také v proměnné x.
- Vzdálenost měřených bodů $[x_i, y_i]$ od funkce $f(x)$: $$d_i^2 = \frac{(x_i - \tilde{x})^2}{\sigma_{x_i}^2} + \frac{\left(y_i - f(\tilde{x})\right)^2}{\sigma_{y_i}^2}$$

Taylor: $f(x_i) + \frac{\partial f(x_i)}{\partial x}(\tilde{x} - x_i)$

- minimum pro $d_i^2 = \frac{\left(y_i - f(x_i)\right)^2}{\sigma_{y_i}^2 + \sigma_{x_i}^2 f'^2(x_i)}$

→ opět minimalizujeme:

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(y_i - f(x_i)\right)^2}{\sigma_{y_i}^2 + \sigma_{x_i}^2 f'^2(x_i)}$$



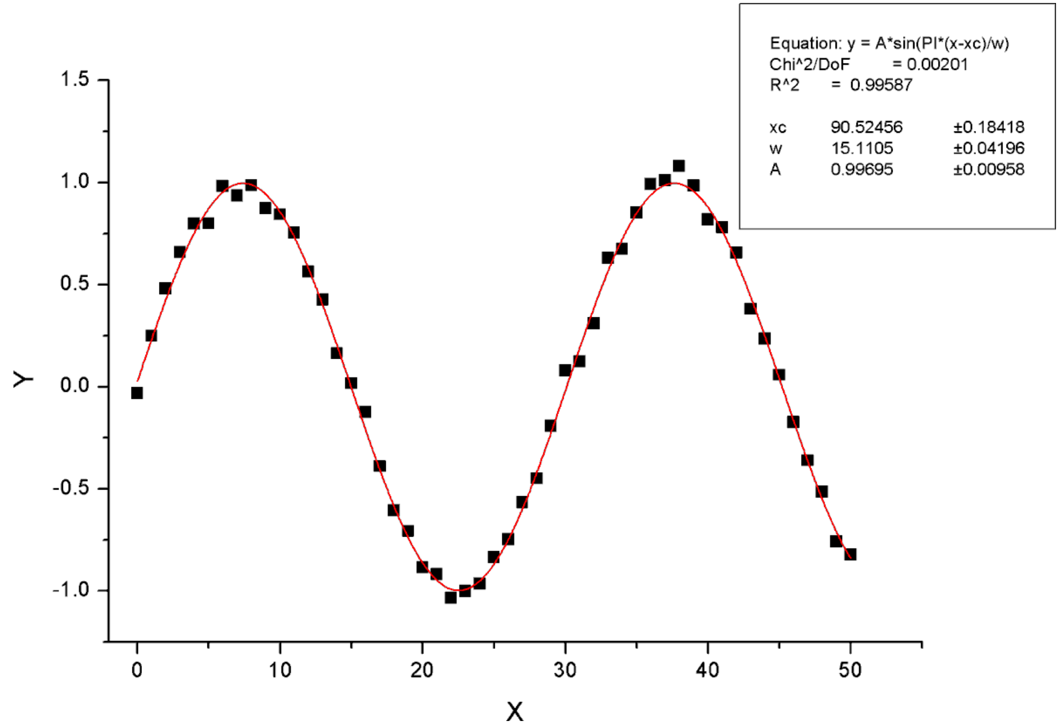
Fitování

- Konstrukce křivky (funkce), která co nejlépe odpovídá naměřeným hodnotám.
  - může podléhat dodatečným podmínkám
- Lineární vs. nelineární regrese

  metoda největšího spádu
  Gaussova-Newtonova metoda
  algoritmus Levenberg–Marquardt
  simplex

- Interpolace a vyhlazování (spline)
- Regresní analýza a extrapolace

- Softwarové nástroje
  - Excel, Matlab, Origin, ...
  - gnuplot, Python, R, ...



Testování hypotéz

Příklad:

Z 30 hodů mincí padl 19x orel a 11x panna. Je mince **poctivá**? (α = 5 %)

Testování hypotéz

Příklad:

Z 30 hodů mincí padl 19x orel a 11x panna. Je mince **poctivá**? (α = 5 %)

nulová hypotéza H_0: mince je poctivá (výsledky se řídí binom. rozdělením s p=½)
alternativní hypotéza H_1: mince není poctivá (nemá binomické rozdělení s p=½)

• spočítáme p-hodnotu: pravděpodobnost, že poctivá mince dá pozorovaný výsledek

$$\sum_{k=19}^{30} B\left(N = 30, k, p = \tfrac{1}{2}\right) = 0{,}100244\ldots$$

• p-hodnota je v našem případě pravděpodobnost, že: padne 19x a více orel, nebo
padne 19x a více panna

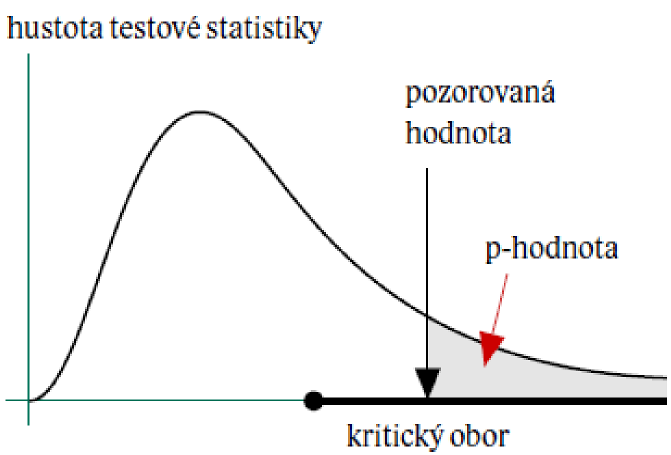
p-hodnota = 2x 0,100244 ~ 0,2

• p-hodnota je větší než hladina významnosti 5%, **hypotézu** tedy **nezamítneme**.

např. pro 21x orel a 9x panna už by p-hodnota byla 0,043 a H_0 bychom zamítli.

Testování hypotéz - pojmy

- **Statistická hypotéza** – testovatelné tvrzení (např. rozdělení zkoumané veličiny, parametry, ...)
- **Test hypotézy** - pravidlo, pomocí kterého hypotézu **zamítneme** nebo **nezamítneme**.
  - obvykle stavíme proti sobě: *nulová hypotéza* H_0 vs. *alternativní hypotéza* H_1
- Chyba: - pokud je platná hypotéza zamítnuta (chyba 1. druhu) α
  - pokud neplatná hypotéza zamítnuta není (chyba 2. druhu) β
  - pravděpodobnost výskytu chyb určuje kvalitu našeho testu.
- **Hladina významnosti α**: pravděpodobnost chyby 1. druhu nepřekročí hodnotu α
- **Síla testu**: $1 - \beta$
- Testovací kritérium - testovací statistika



p-hodnota: jak často nastává situace svědčící proti testované hypotéze.

hypotézu H_0 zamítáme na hladině pravděpodobnosti α, pokud je p-hodnota $< \alpha$

(kritický obor - množina hodnot, pro které test hypotézu zamítá)

χ^2-test

• užitečný při fitování $y = f(x|\alpha_1, \alpha_2, \dots, \alpha_k)$ $x_1, x_2, \dots, xn$

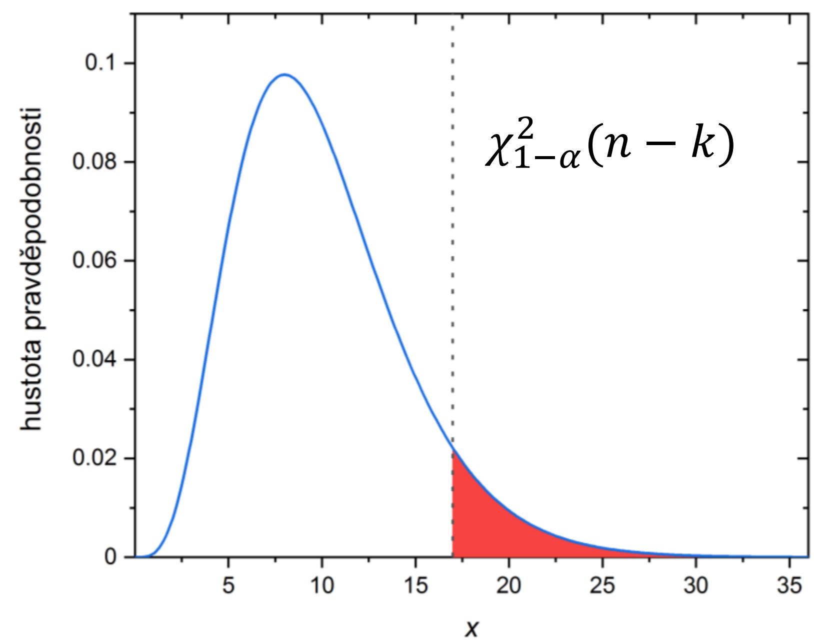
testovací statistika: $y_1, y_2, \dots, yn$

$$X^2 = \sum_{i=1}^{n} \frac{\left(y_i - f(x_i|\alpha_1, \dots, \alpha_k)\right)^2}{\sigma_i^2}$$

$$f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} \exp\left(-\frac{x^2}{2}\right)$$

srovnáváme s χ^2 rozdělením
s $n-k$ stupni volnosti:

Pokud $X^2 > \chi^2_{1-\alpha}(n-k)$,
hypotézu (fit) zamítneme
(na hladině významnosti α)



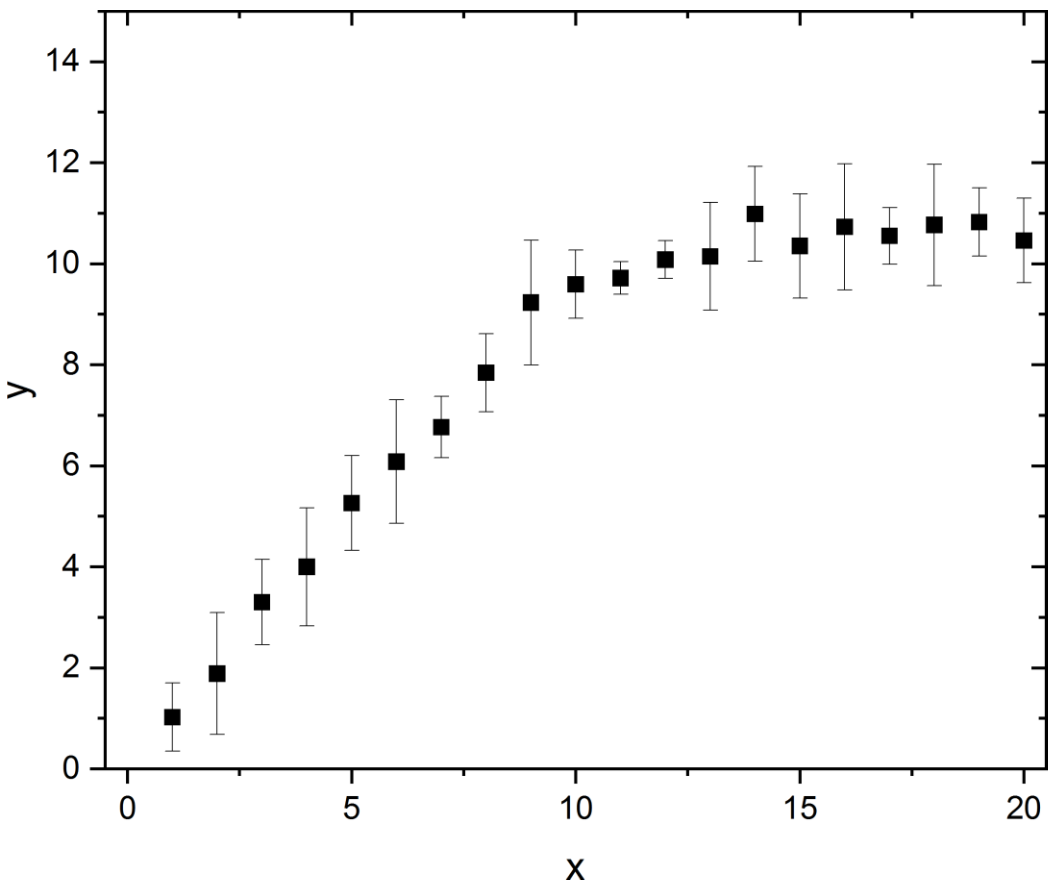
χ^2-test

χ^2 rozdělení s $n - k$ stupni volnosti:

α	0.9	0.7	0.5	0.3	0.2	0.1	**0.05**	**0.02**	**0.01**	**0.001**
n-k										
1	0.02	0.15	0.45	1.07	1.64	2.71	3.84	5.41	6.63	10.83
2	0.21	0.71	1.39	2.41	3.22	4.61	5.99	7.82	9.21	13.82
3	0.58	1.42	2.37	3.66	4.64	6.25	7.81	9.84	11.34	16.27
4	1.06	2.19	3.36	4.88	5.99	7.78	9.49	11.67	13.28	18.47
5	1.61	3.00	4.35	6.06	7.29	9.24	11.07	13.39	15.09	20.52
6	2.20	3.83	5.35	7.23	8.56	10.64	12.59	15.03	16.81	22.46
7	2.83	4.67	6.35	8.38	9.80	12.02	14.07	16.62	18.48	24.32
8	3.49	5.53	7.34	9.52	11.03	13.36	15.51	18.17	20.09	26.12
9	4.17	6.39	8.34	10.66	12.24	14.68	16.92	19.68	21.67	27.88
10	4.87	7.27	9.34	11.78	13.44	15.99	18.31	21.16	23.21	29.59
12	6.30	9.03	11.34	14.01	15.81	18.55	21.03	24.05	26.22	32.91
15	8.55	11.72	14.34	17.32	19.31	22.31	25.00	28.26	30.58	37.70
20	12.44	16.27	19.34	22.77	25.04	28.41	31.41	35.02	37.57	45.31
30	20.60	25.51	29.34	33.53	36.25	40.26	43.77	47.96	50.89	59.70
50	37.69	44.31	49.33	54.72	58.16	63.17	67.50	72.61	76.15	86.66
100	82.36	92.13	99.33	106.91	111.67	118.50	124.34	131.14	135.81	149.45

χ^2-test kvality fitu

n = 20



χ^2-test kvality fitu

$n = 20$

$k = 1, \quad n\text{-}k = 19$

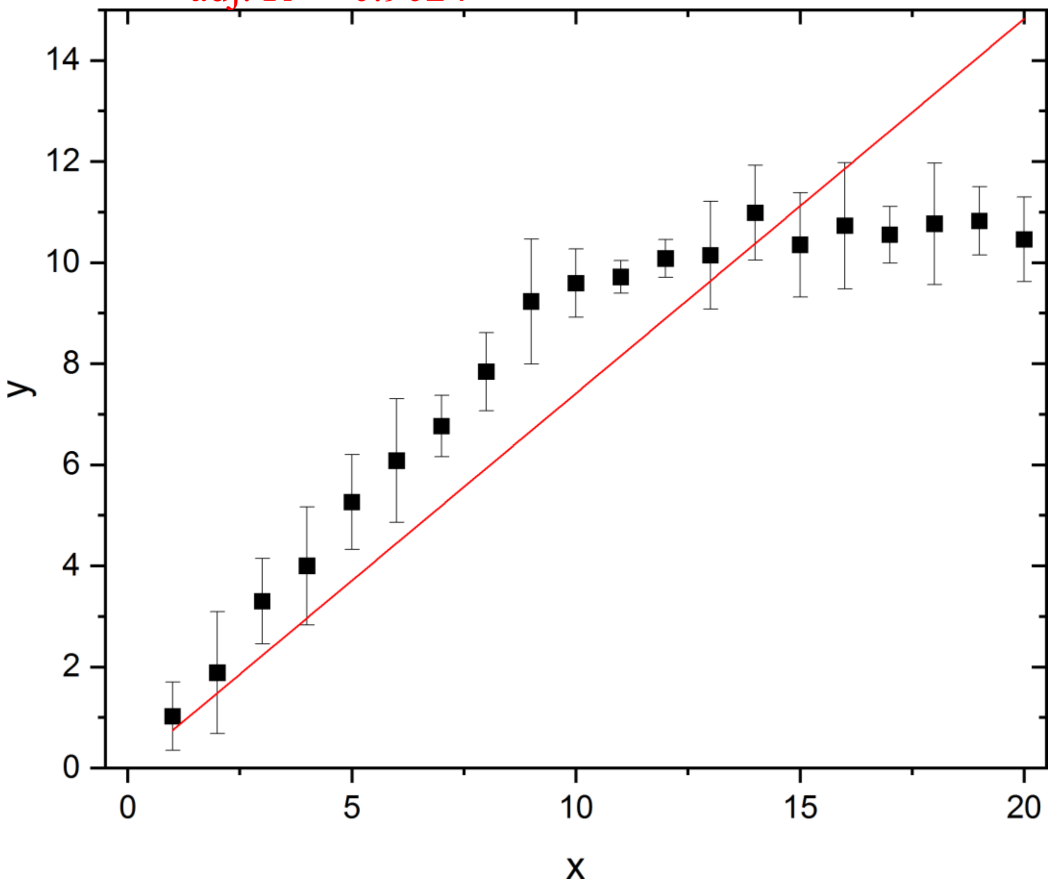
$\chi^2 = 138.77$

$\chi^2 / (n\text{-}k) = 7.304$

$R = 0.9797$

$R^2 = 0.9599$

adj. $R^2 = 0.9024$

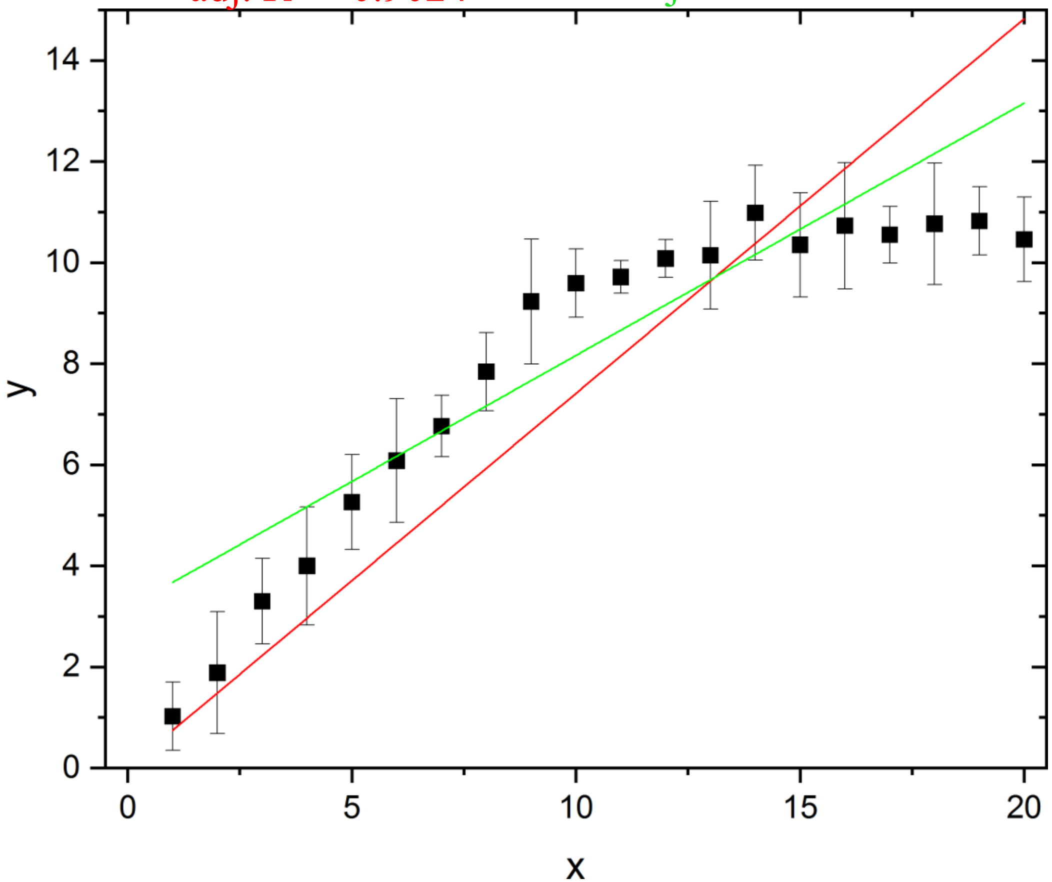


χ^2-test kvality fitu

$n = 20$

$k = 1, \quad n\text{-}k = 19$
$\chi^2 = 138.77$
$\chi^2 / (n\text{-}k) = 7.304$
$R = 0.9797$
$R^2 = 0.9599$
adj. $R^2 = 0.9024$

$k = 2, \quad n\text{-}k = 18$
$\chi^2 = 70.431$
$\chi^2 / (n\text{-}k) = 3.913$
$R = 0.88078$
$R^2 = 0.77577$
adj. $R^2 = 0.76332$



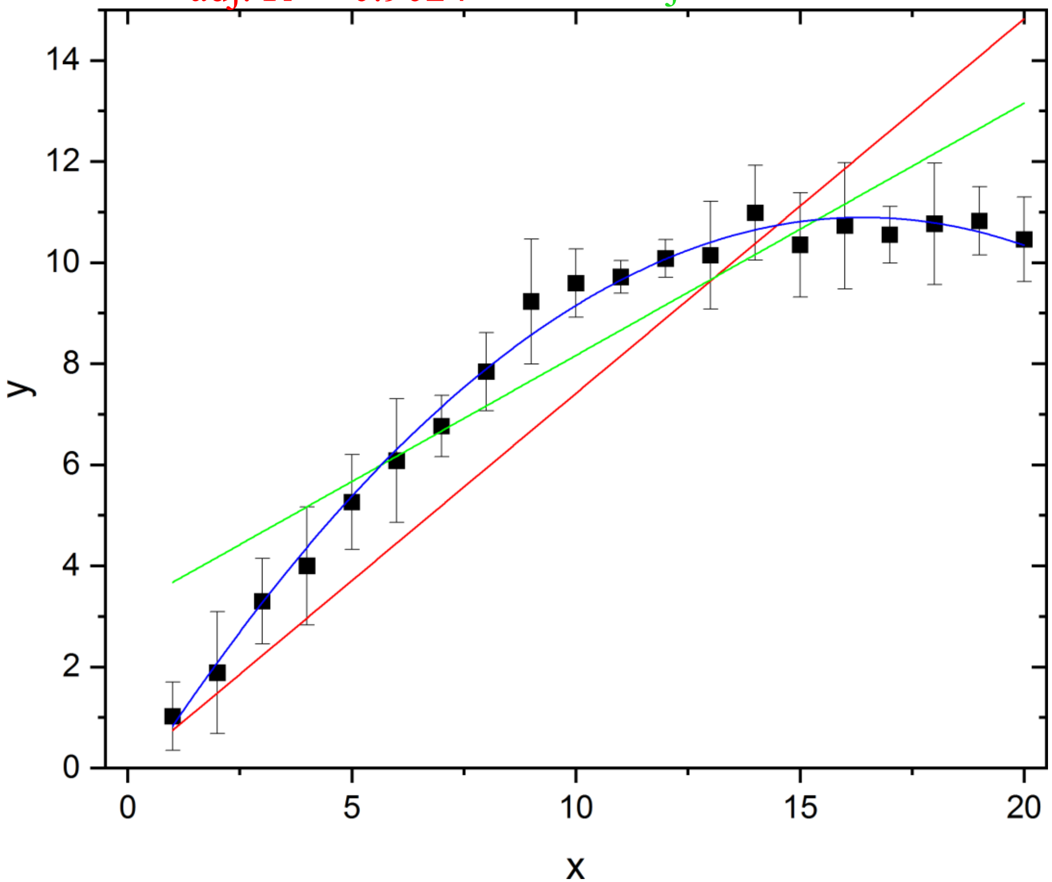
χ^2-test kvality fitu

$n = 20$

$k = 1$, $n\text{-}k = 19$
$\chi^2 = 138.77$
$\chi^2 / (n\text{-}k) = 7.304$
$R = 0.9797$
$R^2 = 0.9599$
adj. $R^2 = 0.9024$

$k = 2$, $n\text{-}k = 18$
$\chi^2 = 70.431$
$\chi^2 / (n\text{-}k) = 3.913$
$R = 0.88078$
$R^2 = 0.77577$
adj. $R^2 = 0.76332$

$k = 3$, $n\text{-}k = 17$
$\chi^2 = 2.2635$
$\chi^2 / (n\text{-}k) = 0.1331$
$R = 0.9964$
$R^2 = 0.9928$
adj. $R^2 = 0.9920$



χ^2-test kvality fitu

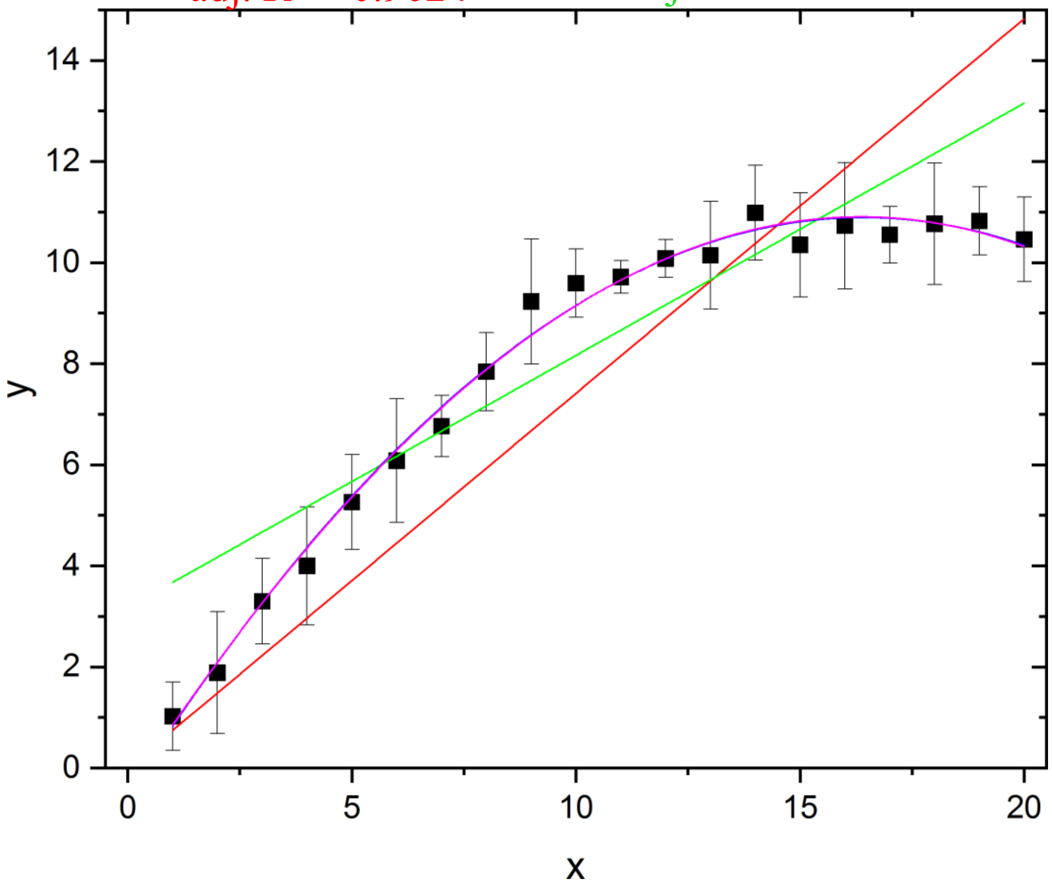
$n = 20$

$k = 1, \quad n\text{-}k = 19$
$\chi^2 = 138.77$
$\chi^2 / (n\text{-}k) = 7.304$
$R = 0.9797$
$R^2 = 0.9599$
adj. $R^2 = 0.9024$

$k = 2, \quad n\text{-}k = 18$
$\chi^2 = 70.431$
$\chi^2 / (n\text{-}k) = 3.913$
$R = 0.88078$
$R^2 = 0.77577$
adj. $R^2 = 0.76332$

$k = 3, \quad n\text{-}k = 17$
$\chi^2 = 2.2635$
$\chi^2 / (n\text{-}k) = 0.1331$
$R = 0.9964$
$R^2 = 0.9928$
adj. $R^2 = 0.9920$

$k = 4, \quad n\text{-}k = 13$
$\chi^2 = 2.25921$
$\chi^2 / (n\text{-}k) = 0.12561$
$R = 0.9964$
$R^2 = 0.9928$
adj. $R^2 = 0.9915$



χ^2-test kvality fitu

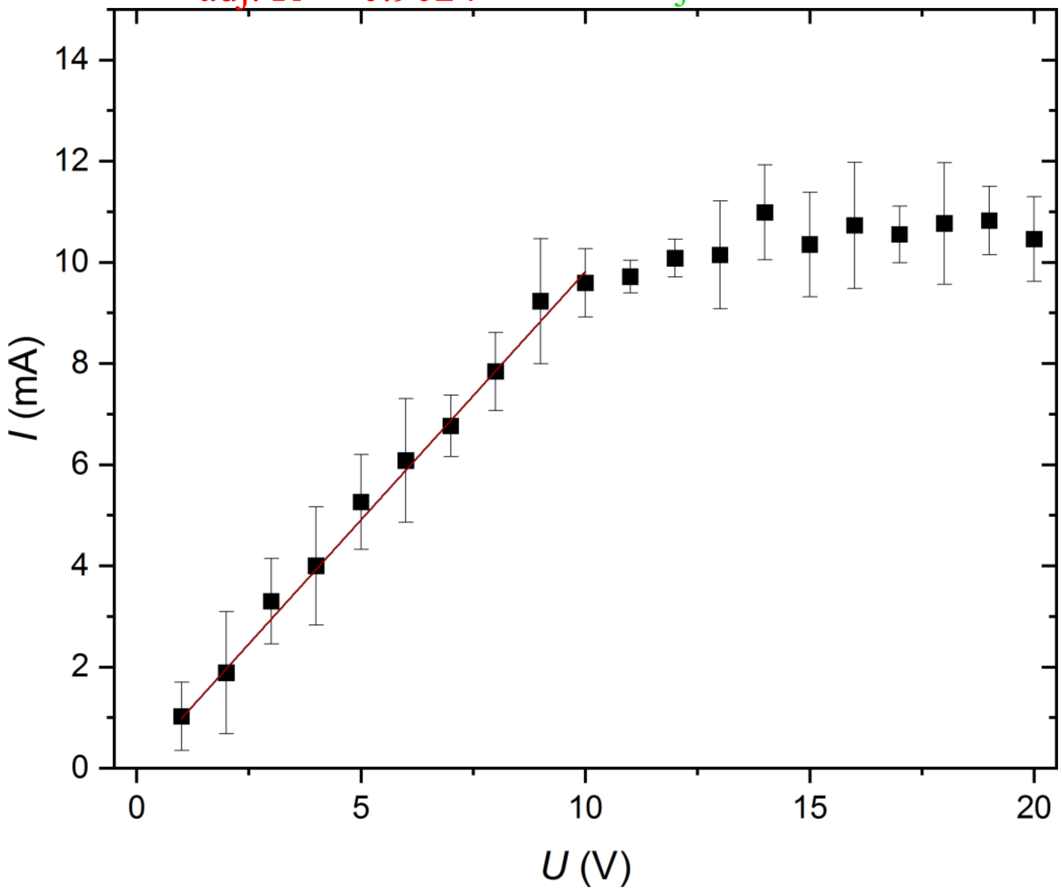
$n = 20$

$k = 1$, $n\text{-}k = 19$
$\chi^2 = 138.77$
$\chi^2 / (n\text{-}k) = 7.304$
$R = 0.9797$
$R^2 = 0.9599$
adj. $R^2 = 0.9024$

$k = 2$, $n\text{-}k = 18$
$\chi^2 = 70.431$
$\chi^2 / (n\text{-}k) = 3.913$
$R = 0.88078$
$R^2 = 0.77577$
adj. $R^2 = 0.76332$

$k = 3$, $n\text{-}k = 17$
$\chi^2 = 2.2635$
$\chi^2 / (n\text{-}k) = 0.1331$
$R = 0.9964$
$R^2 = 0.9928$
adj. $R^2 = 0.9920$

$k = 4$, $n\text{-}k = 13$
$\chi^2 = 2.25921$
$\chi^2 / (n\text{-}k) = 0.12561$
$R = 0.9964$
$R^2 = 0.9928$
adj. $R^2 = 0.9915$



$n = 10$
$k = 1$, $n\text{-}k = 9$
$\chi^2 = 0.59918$
$\chi^2 / (n\text{-}k) = 0.06658$
$R = 0.9995$
$R^2 = 0.9990$
adj. $R^2 = 0.9988$

Residuální analýza, …

Z-test

Pro případy, kdy **známe parametry** μ, σ veličiny x
nebo máme **dostatečný vzorek** ($n \gtrsim 50$)

Testujeme vůči normálnímu rozdělení:

- H_0: $\bar{x} = \mu_0$ (zamítáme, když by $\bar{x} - \mu_0$ bylo příliš velké)

standardizované skóre $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ srovnáváme s $N(0,1)$:

$$p = \int_{-\infty}^{-z} e^{-\frac{x^2}{2}}\,dx + \int_{z}^{\infty} e^{-\frac{x^2}{2}}\,dx$$

Pro naši minci: $z = \frac{19-Np}{\sqrt{Np(1-p)}} = \frac{19-15}{\sqrt{7.5}} \doteq 1.46$

p-hodnota: 0.144 vs. $\alpha = 0.05$

t-test

Pro případy, kdy **neznáme parametry** μ, σ veličiny x
a máme **malý vzorek** $(n \lesssim 50)$

$$f(t) \equiv \frac{x}{y} = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 - \frac{t^2}{n}\right)^{-\frac{n-1}{2}}$$

Testujeme vůči Studentovu t-rozdělení s $n-2$ stupni volnosti:

- H_0: $\bar{x} = \mu_0$ (zamítáme, když by $\bar{x} - \mu_0$ bylo příliš velké)

testovací statistika: $t = \frac{\bar{x} - \mu_0}{\frac{\hat{\sigma}}{\sqrt{n}}}$ $\hat{\sigma}$ je odhad σ ze vzorku x_n

Také se používá pro testování dvou nezávislých vzorků x_n, y_n:

$$t = \frac{\bar{x} - \bar{y} - d}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

Další testování - Benfordův zákon, ...

- Jaké jsou četnosti prvních číslic v datech? → Benfordův zákon

- Jak správně falšovat volby?