Základy biostatistiky

(MD710P09)
ak. rok 2008/2009

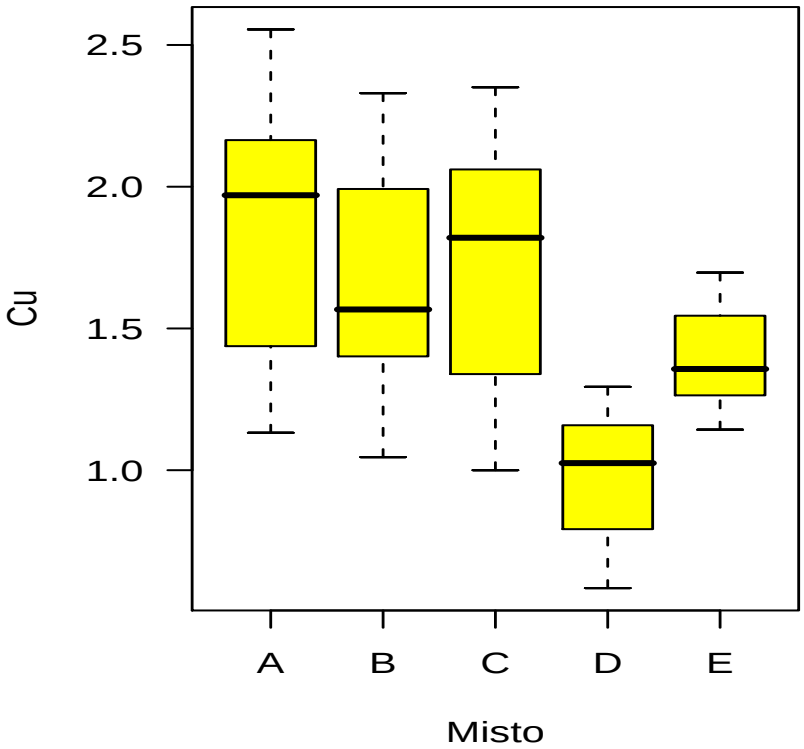
Karel Zvára

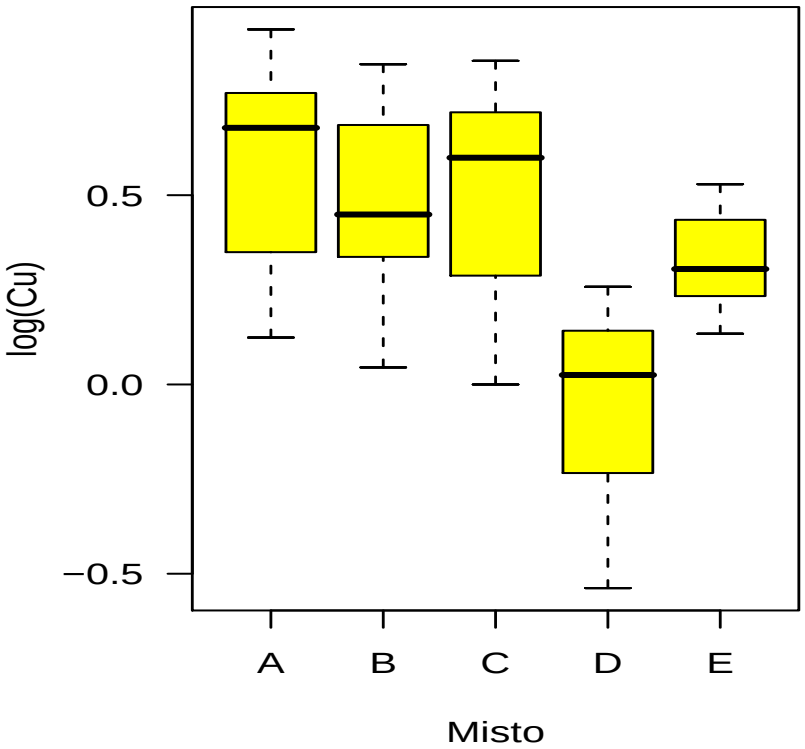
karel.zvara@mff.cuni.cz
http://www.karlin.mff.cuni.cz/∼zvara
katedra pravděpodobnosti a matematické statistiky MFF UK

(naposledy upraveno 20. dubna 2009)

motivační příklad pro analýzu rozptylu (játra):

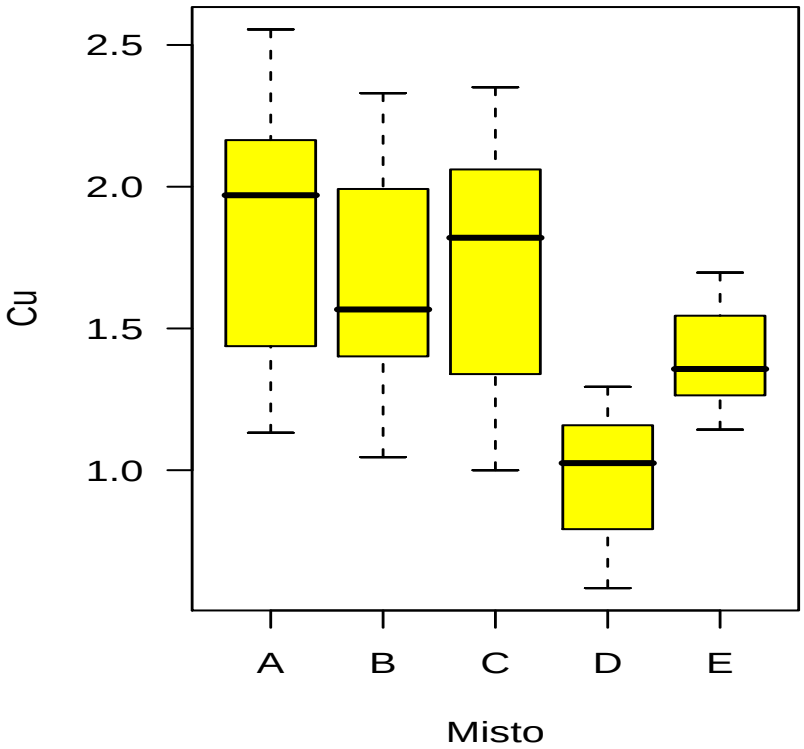
- pět míst na řece, vždy vyloveno po 7 rybách
- zjišťována koncentrace mědi v játrech
- liší se tato místa svým znečištěním?
- logaritmování na pravé straně stabilizuje rozptyl

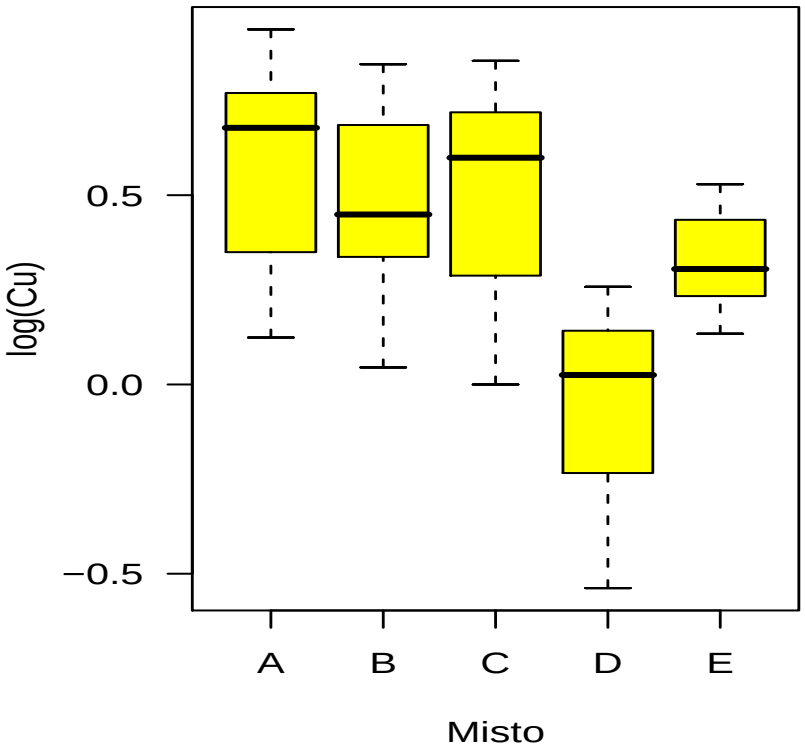




motivační příklad pro analýzu rozptylu (játra):

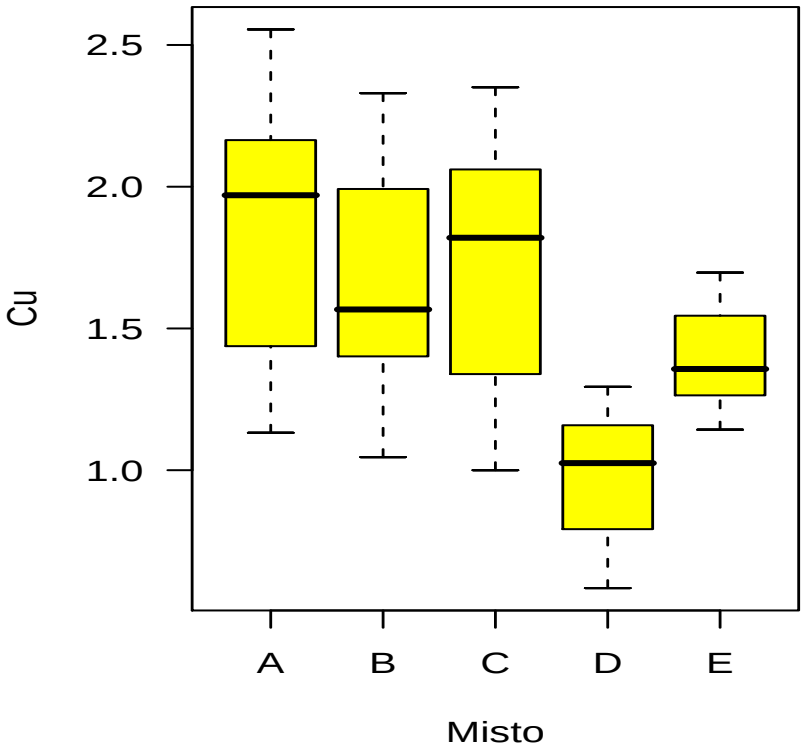
- pět míst na řece, vždy vyloveno po 7 rybách
- zjišťována koncentrace mědi v játrech
- liší se tato místa svým znečištěním?
- logaritmování na pravé straně stabilizuje rozptyl

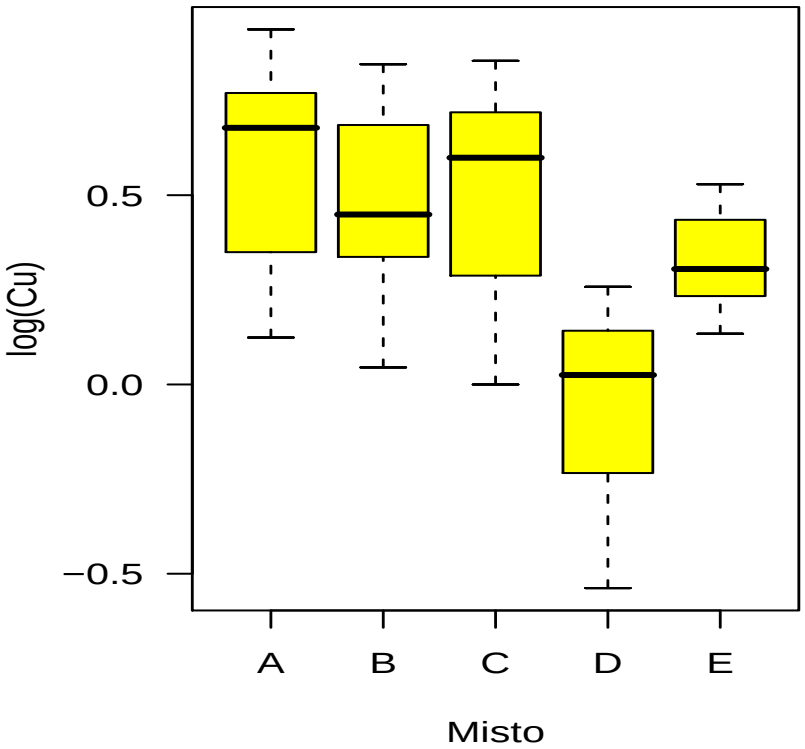




motivační příklad pro analýzu rozptylu (játra):

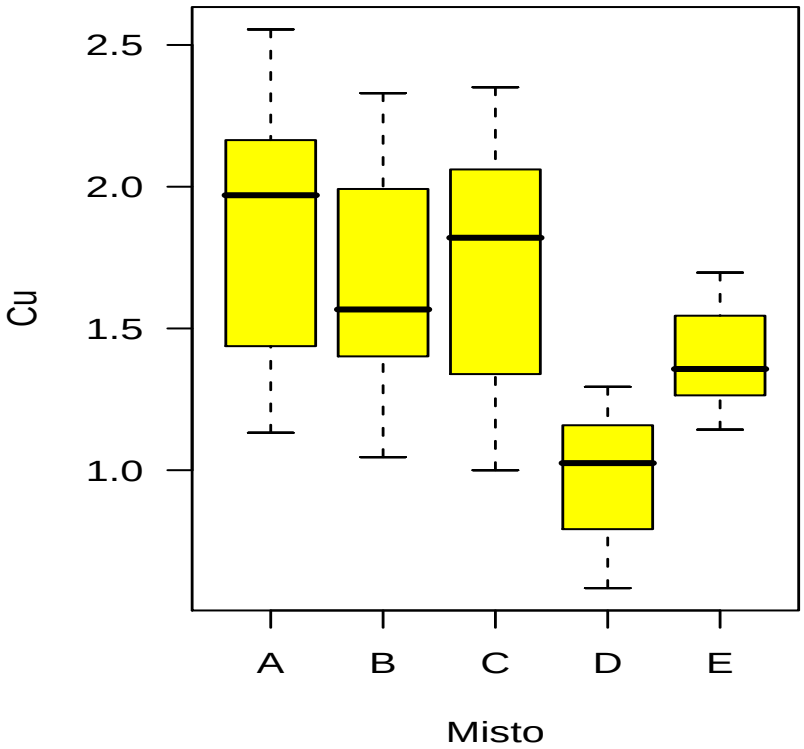
- pět míst na řece, vždy vyloveno po 7 rybách
- zjišťována koncentrace mědi v játrech
- liší se tato místa svým znečištěním?
- logaritmování na pravé straně stabilizuje rozptyl

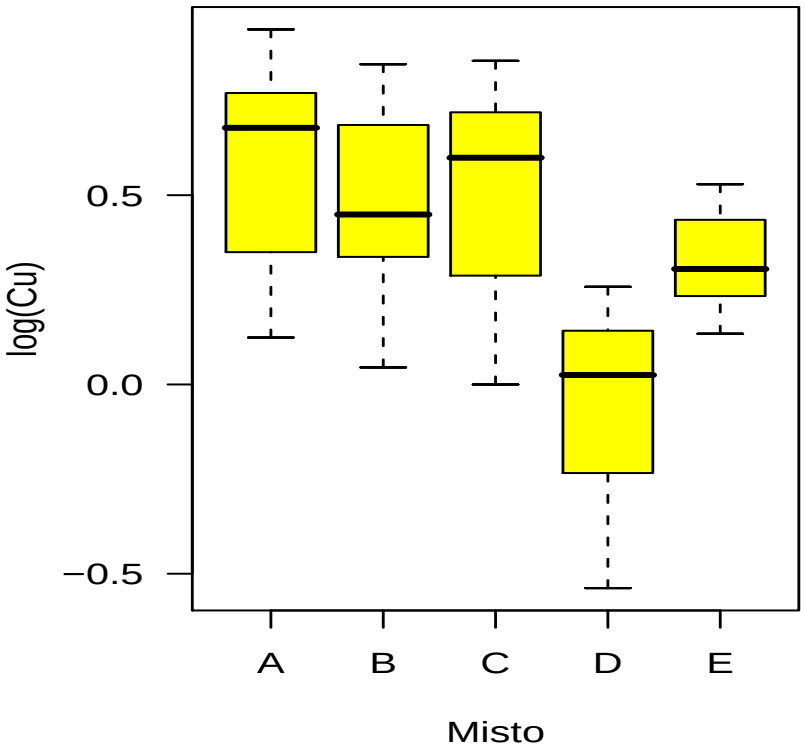




motivační příklad pro analýzu rozptylu (játra):

- pět míst na řece, vždy vyloveno po 7 rybách
- zjišťována koncentrace mědi v játrech
- liší se tato místa svým znečištěním?
- logaritmování na pravé straně stabilizuje rozptyl





analýza rozptylu jednoduchého třídění (ANOVA)

- $Y_{11}, \ldots, Y_{1n_1} \sim \mathrm{N}(\mu_1, \sigma^2)$ (první výběr, průměr $\bar{Y}_{1\bullet}$)
  $Y_{21}, \ldots, Y_{2n_2} \sim \mathrm{N}(\mu_2, \sigma^2)$ (druhý výběr, průměr $\bar{Y}_{2\bullet}$)
  ...
  $Y_{k1}, \ldots, Y_{kn_k} \sim \mathrm{N}(\mu_k, \sigma^2)$ (k-tý výběr, průměr $\bar{Y}_{k\bullet}$)
- **nezávislé** výběry (shodné rozptyly, normální rozdělení)
- $\mathrm{H}_0 : \mu_1 = \mu_2 = \ldots = \mu_k \quad (= \mu)$ H_1 : neplatí H_0
- rozklad součtu čtverců (celkový průměr $\bar{Y}_{\bullet\bullet}$)

$$\sum_{i=1}^{k}\sum_{t=1}^{n_i}(Y_{it} - \bar{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{k} n_i(\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 + \sum_{i=1}^{k}\sum_{t=1}^{n_i}(Y_{it} - \bar{Y}_{i\bullet})^2$$

(celková variabilita) = (variabilita mezi) + (variabilita uvnitř)

$$S_T = S_A + S_e$$

$$f_T = f_A + f_e$$

$$(n-1) = (k-1) + (n-k)$$

analýza rozptylu jednoduchého třídění (ANOVA)

- $Y_{11}, \ldots, Y_{1n_1} \sim \mathrm{N}(\mu_1, \sigma^2)$ (první výběr, průměr $\bar{Y}_{1\bullet}$)
  $Y_{21}, \ldots, Y_{2n_2} \sim \mathrm{N}(\mu_2, \sigma^2)$ (druhý výběr, průměr $\bar{Y}_{2\bullet}$)
  ...
  $Y_{k1}, \ldots, Y_{kn_k} \sim \mathrm{N}(\mu_k, \sigma^2)$ (k-tý výběr, průměr $\bar{Y}_{k\bullet}$)
- **nezávislé** výběry (shodné rozptyly, normální rozdělení)
- $\mathrm{H}_0 : \mu_1 = \mu_2 = \ldots = \mu_k \quad (= \mu)$ H_1 : neplatí H_0
- rozklad součtu čtverců (celkový průměr $\bar{Y}_{\bullet\bullet}$)

$$\sum_{i=1}^{k}\sum_{t=1}^{n_i}(Y_{it} - \bar{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{k} n_i(\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 + \sum_{i=1}^{k}\sum_{t=1}^{n_i}(Y_{it} - \bar{Y}_{i\bullet})^2$$

$$(\text{celková variabilita}) = (\text{variabilita mezi}) + (\text{variabilita uvnitř})$$

$$S_T = S_A + S_e$$

$$f_T = f_A + f_e$$

$$(n-1) = (k-1) + (n-k)$$

analýza rozptylu jednoduchého třídění (ANOVA)

- $Y_{11}, \ldots, Y_{1n_1} \sim \mathrm{N}(\mu_1, \sigma^2)$ (první výběr, průměr $\bar{Y}_{1\bullet}$)
  $Y_{21}, \ldots, Y_{2n_2} \sim \mathrm{N}(\mu_2, \sigma^2)$ (druhý výběr, průměr $\bar{Y}_{2\bullet}$)
  ...
  $Y_{k1}, \ldots, Y_{kn_k} \sim \mathrm{N}(\mu_k, \sigma^2)$ (k-tý výběr, průměr $\bar{Y}_{k\bullet}$)
- **nezávislé** výběry (shodné rozptyly, normální rozdělení)
- $\mathrm{H}_0 : \mu_1 = \mu_2 = \ldots = \mu_k \quad (= \mu)$ H_1 : neplatí H_0
- rozklad součtu čtverců (celkový průměr $\bar{Y}_{\bullet\bullet}$)

$$\sum_{i=1}^{k}\sum_{t=1}^{n_i}(Y_{it} - \bar{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{k} n_i(\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 + \sum_{i=1}^{k}\sum_{t=1}^{n_i}(Y_{it} - \bar{Y}_{i\bullet})^2$$

$$(\text{celková variabilita}) = (\text{variabilita mezi}) + (\text{variabilita uvnitř})$$

$$S_T = S_A + S_e$$

$$f_T = f_A + f_e$$

$$(n-1) = (k-1) + (n-k)$$

analýza rozptylu jednoduchého třídění (ANOVA)

- $Y_{11}, \ldots, Y_{1n_1} \sim \mathrm{N}(\mu_1, \sigma^2)$ (první výběr, průměr $\bar{Y}_{1\bullet}$)
  $Y_{21}, \ldots, Y_{2n_2} \sim \mathrm{N}(\mu_2, \sigma^2)$ (druhý výběr, průměr $\bar{Y}_{2\bullet}$)
  ...
  $Y_{k1}, \ldots, Y_{kn_k} \sim \mathrm{N}(\mu_k, \sigma^2)$ (k-tý výběr, průměr $\bar{Y}_{k\bullet}$)
- **nezávislé** výběry (shodné rozptyly, normální rozdělení)
- $\mathrm{H}_0 : \mu_1 = \mu_2 = \ldots = \mu_k \quad (= \mu)$ H_1 : neplatí H_0
- rozklad součtu čtverců (celkový průměr $\bar{Y}_{\bullet\bullet}$)

$$\sum_{i=1}^{k}\sum_{t=1}^{n_i}(Y_{it} - \bar{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{k} n_i(\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 + \sum_{i=1}^{k}\sum_{t=1}^{n_i}(Y_{it} - \bar{Y}_{i\bullet})^2$$

$$(\text{celková variabilita}) = (\text{variabilita mezi}) + (\text{variabilita uvnitř})$$

$$S_T = S_A + S_e$$

$$f_T = f_A + f_e$$

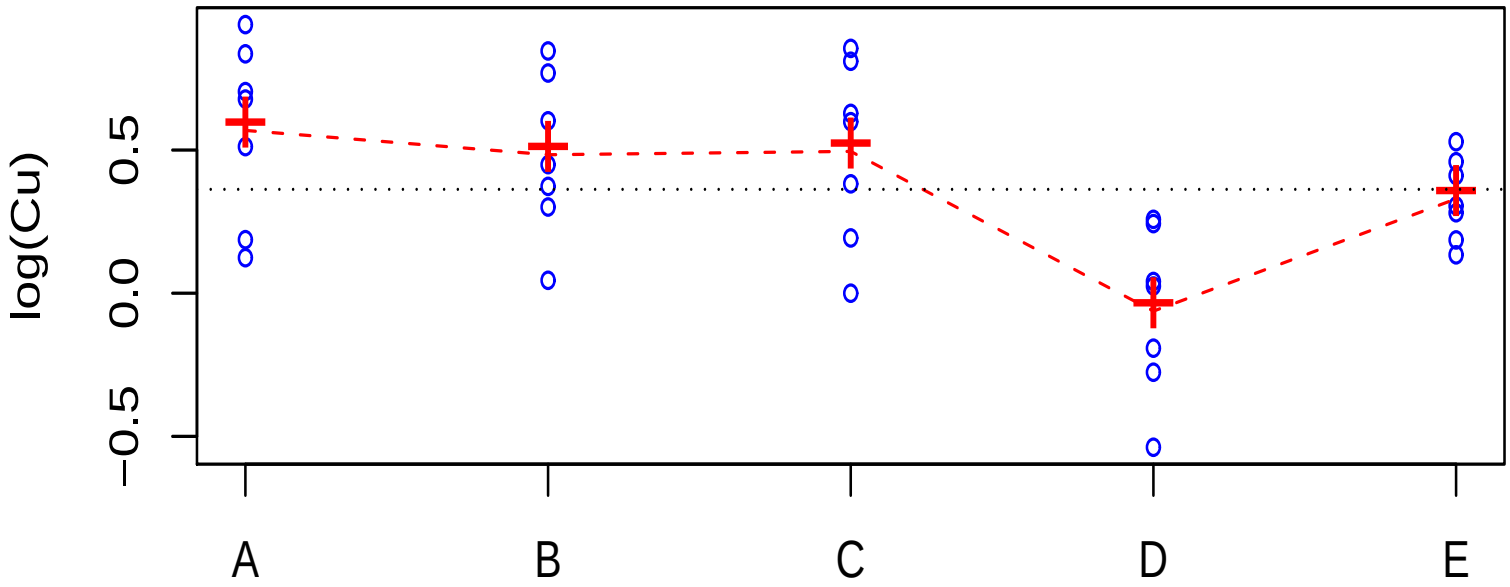
$$(n-1) = (k-1) + (n-k)$$

rozklad součtu čtverců

příklad játra (celkový průměr $\bar{y}_{\bullet\bullet} = 0{,}36$)

(celková variabilita) = (variabilita mezi) + (variabilita uvnitř)

$$\sum_{i=1}^{k}\sum_{t=1}^{n_i}(Y_{it} - \bar{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{k} n_i(\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 + \sum_{i=1}^{k}\sum_{t=1}^{n_i}(Y_{it} - \bar{Y}_{i\bullet})^2$$



tabulka analýzy rozptylu

$$\text{H}_0 \text{ zamítnout, je-li } F_A = \frac{S_A/f_A}{S_e/f_e} \geq F_{f_A,f_e}(\alpha)$$

variabilita	S	f	S/f	F	p
výběry	S_A	$f_A = k-1$	S_A/f_A	F_A	p_A
reziduální	S_e	$f_e = n-k$	S_e/f_e		
celková	S_T	$f_T = n-1$			

- S – součty čtverců, jejich rozklad
- f – počty stupňů volnosti
- S/f – průměrné čtverce
- F – F-statistika
- p – p-hodnota

tabulka analýzy rozptylu

$$\text{H}_0 \text{ zamítnout, je-li } F_A = \frac{S_A/f_A}{S_e/f_e} \geq F_{f_A,f_e}(\alpha)$$

variabilita	S	f	S/f	F	p
výběry	S_A	$f_A = k-1$	S_A/f_A	F_A	p_A
reziduální	S_e	$f_e = n-k$	S_e/f_e		
celková	S_T	$f_T = n-1$			

- S – součty čtverců, jejich rozklad
- f – počty stupňů volnosti
- S/f – průměrné čtverce
- F – F-statistika
- p – p-hodnota

tabulka analýzy rozptylu

$$\text{H}_0 \text{ zamítnout, je-li } F_A = \frac{S_A/f_A}{S_e/f_e} \geq F_{f_A,f_e}(\alpha)$$

variabilita	S	f	S/f	F	p
výběry	S_A	$f_A = k-1$	S_A/f_A	F_A	p_A
reziduální	S_e	$f_e = n-k$	S_e/f_e		
celková	S_T	$f_T = n-1$			

- S – součty čtverců, jejich rozklad
- f – počty stupňů volnosti
- S/f – průměrné čtverce
- F – F-statistika
- p – p-hodnota

tabulka analýzy rozptylu

$$\text{H}_0 \text{ zamítnout, je-li } F_A = \frac{S_A/f_A}{S_e/f_e} \geq F_{f_A,f_e}(\alpha)$$

variabilita	S	f	S/f	F	p
výběry	S_A	$f_A = k-1$	S_A/f_A	F_A	p_A
reziduální	S_e	$f_e = n-k$	S_e/f_e		
celková	S_T	$f_T = n-1$			

- S – součty čtverců, jejich rozklad
- f – počty stupňů volnosti
- S/f – průměrné čtverce
- F – F-statistika
- p – p-hodnota

tabulka analýzy rozptylu

$$\text{H}_0 \text{ zamítnout, je-li } F_A = \frac{S_A/f_A}{S_e/f_e} \geq F_{f_A,f_e}(\alpha)$$

variabilita	S	f	S/f	F	p
výběry	S_A	$f_A = k-1$	S_A/f_A	F_A	p_A
reziduální	S_e	$f_e = n-k$	S_e/f_e		
celková	S_T	$f_T = n-1$			

- S – součty čtverců, jejich rozklad
- f – počty stupňů volnosti
- S/f – průměrné čtverce
- F – F-statistika
- p – p-hodnota

příklad játra

variab.	S	f	S/f	F	p
místa	1,796	4	0,4490	5,862	0,0013
rezid.	2,285	30	0,0762		
celk.	4,081	34			

$F = 5{,}862 > F_{4,30}(0{,}05) = 2{,}690$

na 5% hladině jsme **prokázali rozdíl**

[summary(aov(lnCu∼Misto,data=Med))]

nebo také

[anova(lm(lnCu∼Misto,data=Med))]

varianty zápisu modelu AR jednoduchého třídění

- **model** (měření = úroveň + „chyba“)

$$\begin{aligned} Y_{it} &= \mu_i + E_{it} && 1 \le t \le n_i, \quad 1 \le i \le k \\ &= \mu + (\mu_i - \mu) + E_{it} && E_{it} \text{ nezávislé} \\ &= \mu + \alpha_i + E_{it} && E_{it} \sim \mathrm{N}(0, \sigma^2) \end{aligned}$$

- **reparametrizace** (α_i – **efekty** faktoru A):

$$\sum_{i=1}^{k} \alpha_i = 0$$

- $\mathrm{H}_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k$ (totéž, jako $\mu_1 = \mu_2 = \ldots = \mu_k$)
- pro $k = 2$ je $F_A = T^2$ (vztah s dvouvýběrovým t-testem)

varianty zápisu modelu AR jednoduchého třídění

- **model** (měření = úroveň + „chyba“)

$$\begin{aligned} Y_{it} &= \mu_i + E_{it} \qquad 1 \le t \le n_i, \quad 1 \le i \le k \\ &= \mu + (\mu_i - \mu) + E_{it} \qquad E_{it} \text{ nezávislé} \\ &= \mu + \alpha_i + E_{it} \qquad E_{it} \sim \mathrm{N}(0, \sigma^2) \end{aligned}$$

- **reparametrizace** (α_i – **efekty** faktoru A):

$$\sum_{i=1}^{k} \alpha_i = 0$$

- $\mathrm{H}_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k$ (totéž, jako $\mu_1 = \mu_2 = \ldots = \mu_k$)
- pro $k = 2$ je $F_A = T^2$ (vztah s dvouvýběrovým t-testem)

varianty zápisu modelu AR jednoduchého třídění

- **model** (měření = úroveň + „chyba“)

$$\begin{aligned} Y_{it} &= \mu_i + E_{it} \qquad 1 \le t \le n_i, \quad 1 \le i \le k \\ &= \mu + (\mu_i - \mu) + E_{it} \qquad E_{it} \text{ nezávislé} \\ &= \mu + \alpha_i + E_{it} \qquad E_{it} \sim \mathrm{N}(0, \sigma^2) \end{aligned}$$

- **reparametrizace** (α_i – **efekty** faktoru A):

$$\sum_{i=1}^{k} \alpha_i = 0$$

- $\mathrm{H}_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k$ (totéž, jako $\mu_1 = \mu_2 = \ldots = \mu_k$)
- pro $k = 2$ je $F_A = T^2$ (vztah s dvouvýběrovým t-testem)

varianty zápisu modelu AR jednoduchého třídění

- **model** (měření = úroveň + „chyba“)

$$\begin{aligned} Y_{it} &= \mu_i + E_{it} \qquad 1 \le t \le n_i, \quad 1 \le i \le k \\ &= \mu + (\mu_i - \mu) + E_{it} \qquad E_{it} \text{ nezávislé} \\ &= \mu + \alpha_i + E_{it} \qquad E_{it} \sim \mathrm{N}(0, \sigma^2) \end{aligned}$$

- **reparametrizace** (α_i – **efekty** faktoru A):

$$\sum_{i=1}^{k} \alpha_i = 0$$

- $\mathrm{H}_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k$ (totéž, jako $\mu_1 = \mu_2 = \ldots = \mu_k$)
- pro $k = 2$ je $F_A = T^2$ (vztah s dvouvýběrovým t-testem)

ověření předpokladů

- **nezávislost**: dáno organizací (plánem) pokusu
  předpoklad nelze vynechat či nahradit
- **shoda rozptylů**: (vyvážený model málo citlivý na neshodu)
  - Leveneův test
    (vlastně ANOVA s $|Y_{it} - \mathrm{med}_t\, Y_{it}|$)
    $p = 64{,}8\ \%$ [levene.test(lnCu,Misto)]
  - Bartlettův test
    (citlivý na splnění předpokladu o normálním rozdělení)
    $p = 45{,}3\ \%$ [bartlett.test(lnCu,Misto)]
- **normální rozdělení**: (vyvážený model málo citlivý)
  test normality nutno uplatnit na rezidua $Y_{it} - \bar{Y}_{i\bullet}$
  $p = 6{,}8\ \%$
  [shapiro.test(resid(aov(lnCu Misto)))]
  nebo [shapiro.test(resid(lm(lnCu∼Misto)))]

ověření předpokladů

- **nezávislost**: dáno organizací (plánem) pokusu
  předpoklad nelze vynechat či nahradit
- **shoda rozptylů**: (vyvážený model málo citlivý na neshodu)
  - Leveneův test
    (vlastně ANOVA s $|Y_{it} - \mathrm{med}_t\, Y_{it}|$)
    $p = 64{,}8\ \%$ [levene.test(lnCu,Misto)]
  - Bartlettův test
    (citlivý na splnění předpokladu o normálním rozdělení)
    $p = 45{,}3\ \%$ [bartlett.test(lnCu,Misto)]
- **normální rozdělení**: (vyvážený model málo citlivý)
  test normality nutno uplatnit na rezidua $Y_{it} - \bar{Y}_{i\bullet}$
  $p = 6{,}8\ \%$
  [shapiro.test(resid(aov(lnCu Misto)))]
  nebo [shapiro.test(resid(lm(lnCu∼Misto)))]

ověření předpokladů

- **nezávislost**: dáno organizací (plánem) pokusu
  předpoklad nelze vynechat či nahradit
- **shoda rozptylů**: (vyvážený model málo citlivý na neshodu)
  - Leveneův test
    (vlastně ANOVA s $|Y_{it} - \mathrm{med}_t\, Y_{it}|$)
    $p = 64{,}8\ \%$ [levene.test(lnCu,Misto)]
  - Bartlettův test
    (citlivý na splnění předpokladu o normálním rozdělení)
    $p = 45{,}3\ \%$ [bartlett.test(lnCu,Misto)]
- **normální rozdělení**: (vyvážený model málo citlivý)
  test normality nutno uplatnit na rezidua $Y_{it} - \bar{Y}_{i\bullet}$
  $p = 6{,}8\ \%$
  [shapiro.test(resid(aov(lnCu Misto)))]
  nebo [shapiro.test(resid(lm(lnCu∼Misto)))]

ověření předpokladů

- **nezávislost**: dáno organizací (plánem) pokusu
  předpoklad nelze vynechat či nahradit
- **shoda rozptylů**: (vyvážený model málo citlivý na neshodu)
  - Leveneův test
    (vlastně ANOVA s $|Y_{it} - \mathrm{med}_t\, Y_{it}|$)
    $p = 64{,}8\ \%$ [levene.test(lnCu,Misto)]
  - Bartlettův test
    (citlivý na splnění předpokladu o normálním rozdělení)
    $p = 45{,}3\ \%$ [bartlett.test(lnCu,Misto)]
- **normální rozdělení**: (vyvážený model málo citlivý)
  test normality nutno uplatnit na rezidua $Y_{it} - \bar{Y}_{i\bullet}$
  $p = 6{,}8\ \%$
  [shapiro.test(resid(aov(lnCu Misto)))]
  nebo [shapiro.test(resid(lm(lnCu∼Misto)))]

ověření předpokladů

- **nezávislost**: dáno organizací (plánem) pokusu
  předpoklad nelze vynechat či nahradit
- **shoda rozptylů**: (vyvážený model málo citlivý na neshodu)
  - Leveneův test
    (vlastně ANOVA s $|Y_{it} - \mathrm{med}_t\, Y_{it}|$)
    $p = 64{,}8\ \%$ [levene.test(lnCu,Misto)]
  - Bartlettův test
    (citlivý na splnění předpokladu o normálním rozdělení)
    $p = 45{,}3\ \%$ [bartlett.test(lnCu,Misto)]
- **normální rozdělení**: (vyvážený model málo citlivý)
  test normality nutno uplatnit na rezidua $Y_{it} - \bar{Y}_{i\bullet}$
  $p = 6{,}8\ \%$

  [shapiro.test(resid(aov(lnCu Misto)))]

  nebo [shapiro.test(resid(lm(lnCu~Misto)))]

mnohonásobná srovnání
(Tukeyův test, Kramerova verze)

- nutnost zachovat zvolenou hladinu testu i při současném rozhodování o řadě hypotéz
  (např. že $\mu_1 = \mu_2, \quad \mu_1 = \mu_3, \quad \mu_2 = \mu_3, \quad \ldots$)
- které dvojice úrovní faktoru (stř. hodnoty μ_i resp. efekty α_i) se liší?

$$|\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}| \geq q_{k,n-k}(\alpha)\sqrt{\frac{S^2}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

kde $q_{k,n-k}(\alpha)$ je tabelovaná kritická hodnota

$$S^2 = \frac{S_e}{f_e} = \frac{\sum\sum(Y_{it} - \bar{Y}_{i\bullet})^2}{n-k}$$

mnohonásobná srovnání
(Tukeyův test, Kramerova verze)

- nutnost zachovat zvolenou hladinu testu i při současném rozhodování o řadě hypotéz
  (např. že $\mu_1 = \mu_2, \quad \mu_1 = \mu_3, \quad \mu_2 = \mu_3, \quad \ldots$)
- které dvojice úrovní faktoru (stř. hodnoty μ_i resp. efekty α_i) se liší?

$$|\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}| \geq q_{k,n-k}(\alpha)\sqrt{\frac{S^2}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

kde $q_{k,n-k}(\alpha)$ je tabelovaná kritická hodnota

$$S^2 = \frac{S_e}{f_e} = \frac{\sum\sum(Y_{it} - \bar{Y}_{i\bullet})^2}{n-k}$$

příklad játra

místo	počet	průměr	efekt	směr. odchylka
A	7	0,568	0,206	0,312
B	7	0,484	0,121	0,279
C	7	0,495	0,133	0,318
D	7	-0,063	-0,426	0,290
E	7	0,329	-0,034	0,144
celkem	35	0,363	0,000	0,104

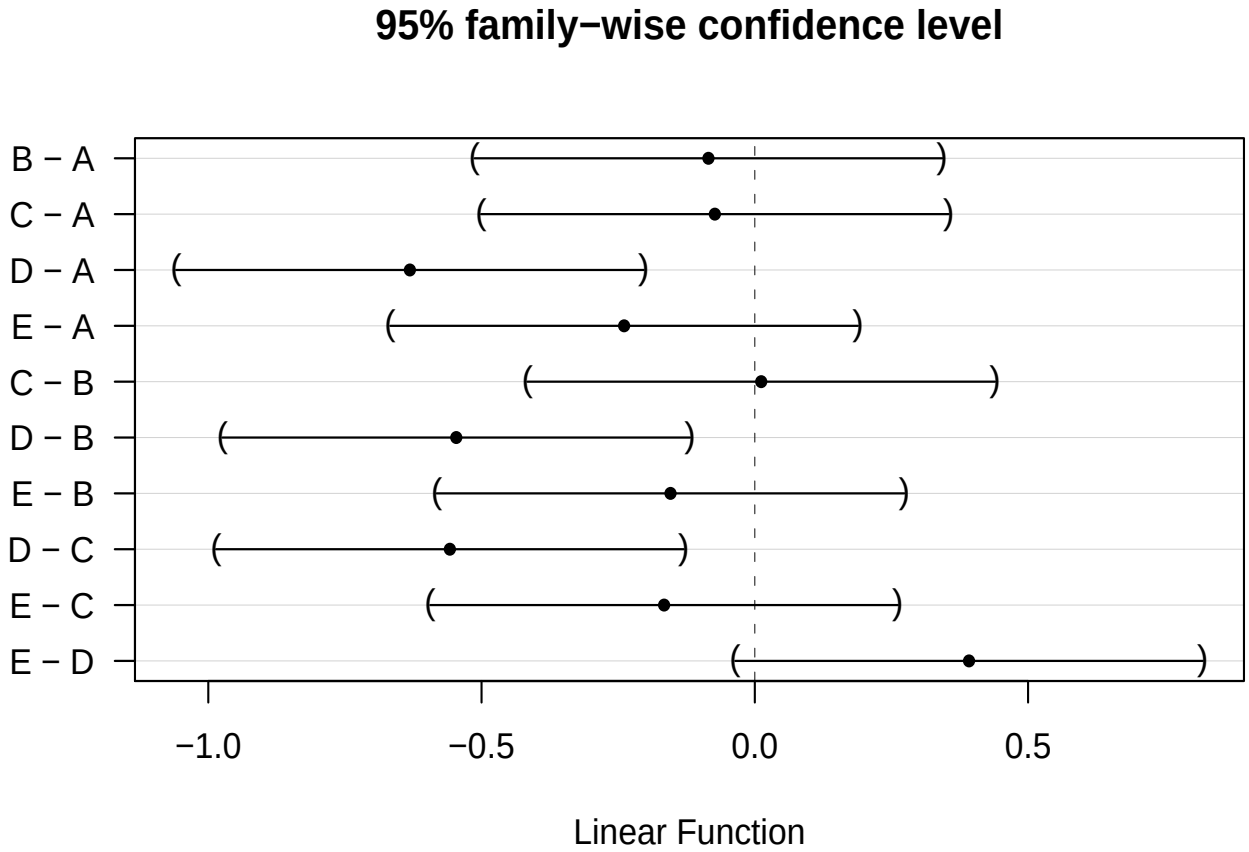
$$q_{5,30}(0{,}05)\sqrt{\frac{0{,}0762}{2}\left(\frac{1}{7}+\frac{1}{7}\right)} = 4{,}10 \cdot 0{,}104 = 0{,}428$$

$-0{,}063 + 0{,}428 = 0{,}365 \Rightarrow$ na 5% hladině se místa D s nejmenším průměrem liší všechna místa s průměry aspoň 0,365, tedy místa A, B, C, nikoliv E

[TukeyHSD(aov(lnCu∼Misto,data=Med))]

příklad játra

funkce [TukeyHSD(aov(lnCu∼Misto,data=Med))]
dá tabulku porovnání všech dvojic
pomocí knihovny Rcmdr dostaneme také graf



Kruskalův-Wallisův text
(neparametrický test)

- zobecnění dvouvýběrového Wilcoxonova testu (použije opět pořadí místo původních hodnot)
- předpoklady:
  - k nezávislých výběrů
  - spojitá rozdělení
- H_0: rozdělení jsou stejná (tedy i mediány jsou stejné)
- T_i - součet pořadí v i-tém výběru

$$Q = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{T_i^2}{n_i} - 3(n+1)$$

H_0 se zamítá při $Q \geq \chi^2_{k-1}(\alpha)$
(velká variabilita průměrných pořadí)

Kruskalův-Wallisův text
(neparametrický test)

- zobecnění dvouvýběrového Wilcoxonova testu (použije opět pořadí místo původních hodnot)
- předpoklady:
  - k nezávislých výběrů
  - spojitá rozdělení
- H_0: rozdělení jsou stejná (tedy i mediány jsou stejné)
- T_i - součet pořadí v i-tém výběru

$$Q = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{T_i^2}{n_i} - 3(n+1)$$

H_0 se zamítá při $Q \geq \chi^2_{k-1}(\alpha)$
(velká variabilita průměrných pořadí)

Kruskalův-Wallisův text
(neparametrický test)

- zobecnění dvouvýběrového Wilcoxonova testu (použije opět pořadí místo původních hodnot)
- předpoklady:
  - k nezávislých výběrů
  - spojitá rozdělení
- H_0: rozdělení jsou stejná (tedy i mediány jsou stejné)
- T_i - součet pořadí v i-tém výběru

$$Q = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{T_i^2}{n_i} - 3(n+1)$$

H_0 se zamítá při $Q \geq \chi^2_{k-1}(\alpha)$
(velká variabilita průměrných pořadí)

Kruskalův-Wallisův text
(neparametrický test)

- zobecnění dvouvýběrového Wilcoxonova testu (použije opět pořadí místo původních hodnot)
- předpoklady:
  - k nezávislých výběrů
  - spojitá rozdělení
- H_0: rozdělení jsou stejná (tedy i mediány jsou stejné)
- T_i - součet pořadí v i-tém výběru

$$Q = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{T_i^2}{n_i} - 3(n+1)$$

H_0 se zamítá při $Q \geq \chi^2_{k-1}(\alpha)$
(velká variabilita průměrných pořadí)

Kruskalův-Wallisův text
(neparametrický test)

- zobecnění dvouvýběrového Wilcoxonova testu (použije opět pořadí místo původních hodnot)
- předpoklady:
  - k nezávislých výběrů
  - spojitá rozdělení
- H_0: rozdělení jsou stejná (tedy i mediány jsou stejné)
- T_i - součet pořadí v i-tém výběru

$$Q = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{T_i^2}{n_i} - 3(n+1)$$

H_0 se zamítá při $Q \geq \chi^2_{k-1}(\alpha)$
(velká variabilita průměrných pořadí)

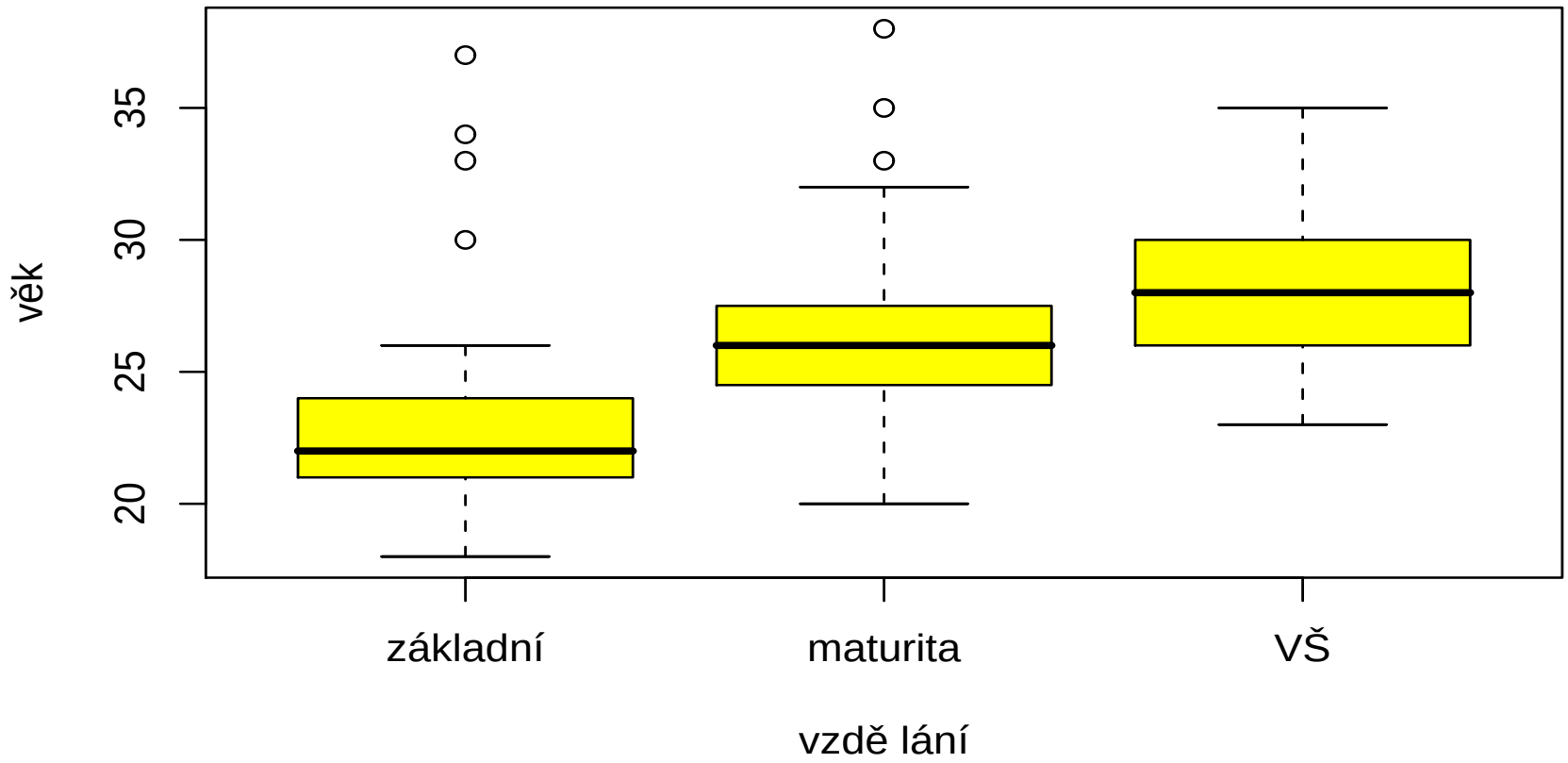
Kruskalův-Wallisův text
(neparametrický test)

- zobecnění dvouvýběrového Wilcoxonova testu
  (použije opět pořadí místo původních hodnot)
- předpoklady:
  - k nezávislých výběrů
  - spojitá rozdělení
- H_0: rozdělení jsou stejná (tedy i mediány jsou stejné)
- T_i - součet pořadí v i-tém výběru

$$Q = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{T_i^2}{n_i} - 3(n+1)$$

H_0 se zamítá při $Q \geq \chi^2_{k-1}(\alpha)$
(velká variabilita průměrných pořadí)

příklad kojení – věk matek podle vzdělání



je patrná nesymetrie, zejména u základního vzdělání

příklad kojení – věk matek podle vzdělání

vzdělání	n_i	průměrný věk	střední chyba	součet pořadí	průměrné pořadí
základní	34	23,412	0,638	1025	30,15
maturita	47	26,278	0,543	2618	55,70
VŠ	18	28,500	0,877	1307	72,61
celk.	99	25,697		4950	50,00

$$Q = \frac{12}{99 \cdot 100}\left(\frac{1025^2}{34} + \frac{2618^2}{47} + \frac{1307^2}{18}\right) - 3 \cdot 100 = 29{,}25$$

$$\chi^2_2(0{,}05) = 5{,}99 \qquad p < 0{,}0001$$

[kruskal.test(vek.m∼Vzdelani,data=Kojeni)]