Základy biostatistiky

(MD710P09)
ak. rok 2007/2008

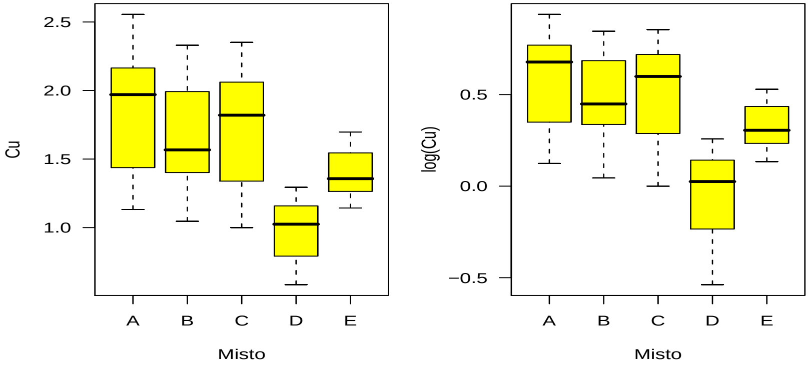
Karel Zvára

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

(naposledy upraveno 10. dubna 2008)

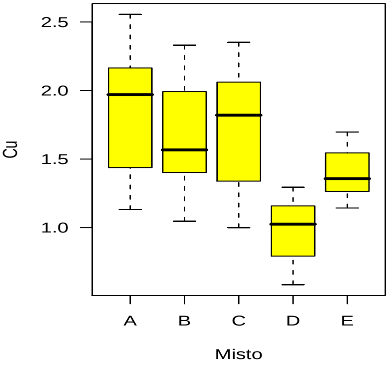
motivační příklad pro analýzu rozptylu (játra):

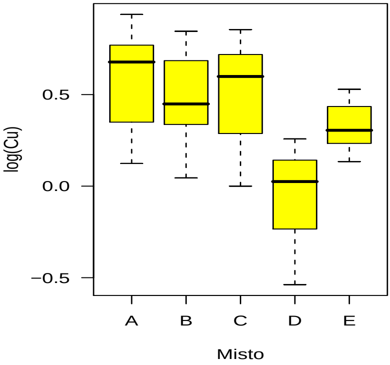
- pět míst na řece, vždy vyloveno po 7 rybách
- zjišťována koncentrace mědi v játrech
- liší se tato místa svým znečištěním?
- logaritmování na pravé straně stabilizuje rozptyl



motivační příklad pro analýzu rozptylu (játra):

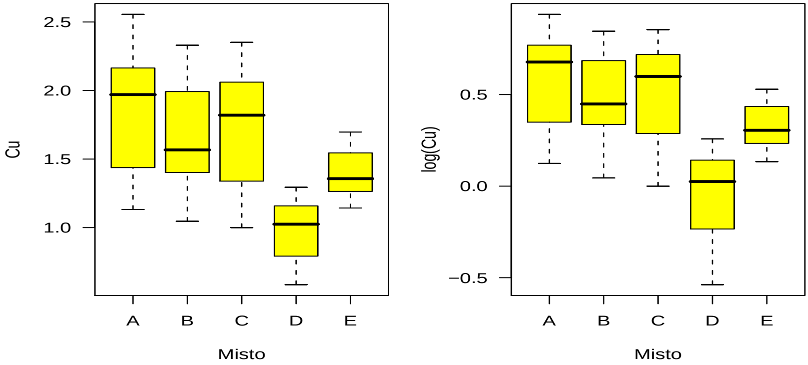
- pět míst na řece, vždy vyloveno po 7 rybách
- zjišťována koncentrace mědi v játrech
- liší se tato místa svým znečištěním?
- logaritmování na pravé straně stabilizuje rozptyl





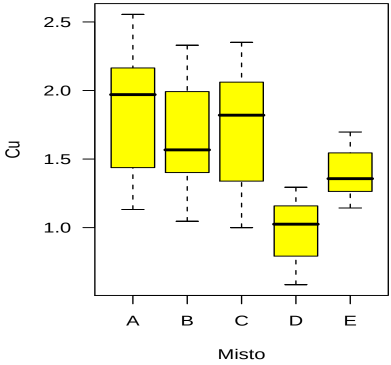
motivační příklad pro analýzu rozptylu (játra):

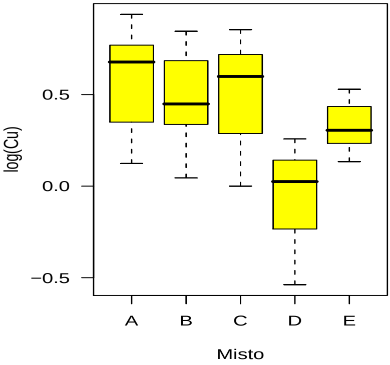
- pět míst na řece, vždy vyloveno po 7 rybách
- zjišťována koncentrace mědi v játrech
- liší se tato místa svým znečištěním?
- logaritmování na pravé straně stabilizuje rozptyl



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

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

(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)

- $Y_{11}, \dots, Y_{1n_1} \sim \mathrm{N}(\mu_1, \sigma^2)$ (první výběr, průměr $\bar{Y}_{1\bullet}$)
  $Y_{21}, \dots, Y_{2n_2} \sim \mathrm{N}(\mu_2, \sigma^2)$ (druhý výběr, průměr $\bar{Y}_{2\bullet}$)
  ...
  $Y_{k1}, \dots, Y_{kn_k} \sim \mathrm{N}(\mu_k, \sigma^2)$ (k-tý výběr, průměr $\bar{Y}_{k\bullet}$)
- **nezávislé** výběry (shodné rozptyly, normální rozdělení)
- $\mathrm{H}_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad (= \mu) \qquad \mathrm{H}_1$: neplatí H_0
- rozklad součtu čtverců (celkový průměr $\bar{Y}_{\bullet\bullet}$)

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

(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)
- $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k \quad (= \mu)$ $\qquad H_1$: neplatí H_0
- rozklad součtu čtverců (celkový průměr $\bar{Y}_{\bullet\bullet}$)

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

(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)
- $\mathrm{H}_0 : \mu_1 = \mu_2 = \ldots = \mu_k \quad (= \mu)$ $\quad \mathrm{H}_1$: neplatí H_0
- rozklad součtu čtverců (celkový průměr $\bar{Y}_{\bullet\bullet}$)

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

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

$$S_T = S_A + S_e$$

$$f_T = f_A + f_e$$

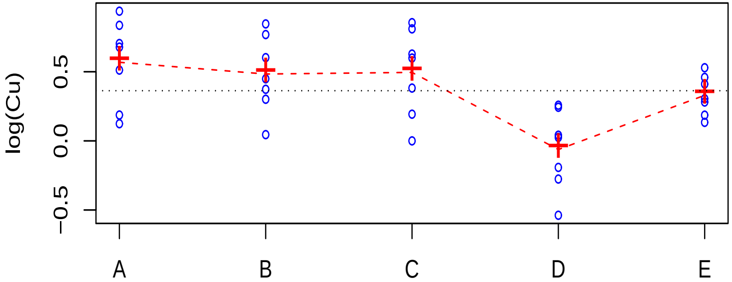
$$(n-1) = (k-1) + (n-k)$$

rozklad součtu čtverců

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

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

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



tabulka analýzy rozptylu

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

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

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

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

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

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

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

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

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

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):

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

- $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k$ (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“)

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

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

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

- $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k$ (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“)

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

- **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):

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

- $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_k$ (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)
  - 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)
  - Leveneův test
    (vlastně ANOVA s $|Y_{it} - \text{med}_t Y_{it}|$)
    $p = 64{,}8\ \%$ [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)
  - Leveneův test
    (vlastně ANOVA s $|Y_{it} - \text{med}_t Y_{it}|$)
    $p = 64{,}8\ \%$ [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)
  - 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)
  - 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
- 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 a

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

mnohonásobná srovnání

(Tukeyův test, Kramerova verze)

- nutnost zachovat zvolenou hladinu testu
- 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 a

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

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

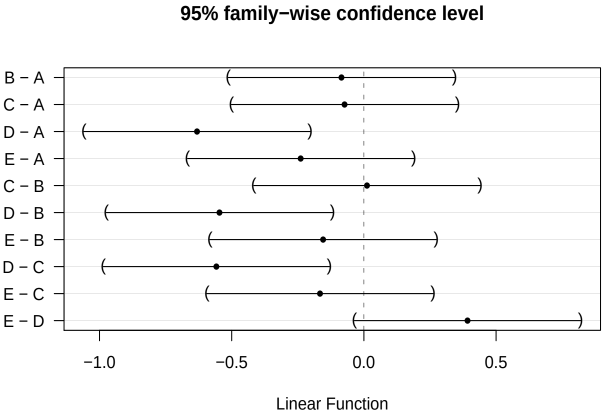
$$q_{5,30}(0{,}05)\sqrt{\frac{0{,}0762}{2}\left(\frac{1}{7}+\frac{1}{7}\right)} = 4{,}10 \cdot 0{,}104 = 0{,}428$$

$-0{,}063 + 0{,}428 = 0{,}365 \Rightarrow$ 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 (použijí se 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 (použijí se 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 (použijí se 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 (použijí se 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 (použijí se 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)

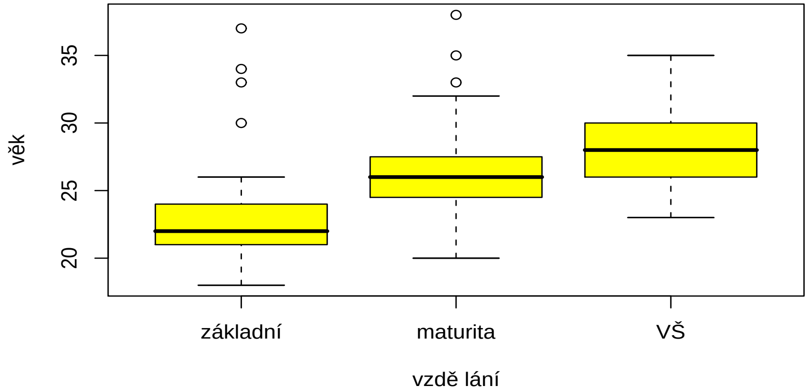
Kruskalův-Wallisův text
(neparametrický test)

- zobecnění dvouvýběrového Wilcoxonova testu (použijí se 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)]

náhodné bloky

- zobecnění párových testů na r-tice
- **náhodný blok**
  - homogenní skupina r objektů
  - počet objektů ve skupině = počet ošetření (nebo jeho násobek)
  - ošetření se přiřadí uvnitř bloku **náhodně** (každému ošetření stejný počet objektů)
- bloky – náhodné efekty $A_i \sim \mathrm{N}(0, \sigma_A^2)$ (vliv bloku)
  ošetření – pevné efekty β_j $\left(\sum_{j=1}^{r} \beta_j = 0\right)$ (vliv ošetření)

$$Y_{ij} = \mu + A_i + \beta_j + E_{ij}, \quad E_{ij} \sim \mathrm{N}(0, \sigma^2), \quad j = 1, \ldots, r, \quad i = 1, \ldots, k$$

předpokládá se **aditivní** vliv, symbolicky zapisovaný $A + B$

náhodné bloky

- zobecnění párových testů na r-tice
- **náhodný blok**
  - homogenní skupina r objektů
  - počet objektů ve skupině = počet ošetření (nebo jeho násobek)
  - ošetření se přiřadí uvnitř bloku **náhodně** (každému ošetření stejný počet objektů)
- bloky – náhodné efekty $A_i \sim \mathrm{N}(0, \sigma_A^2)$ (vliv bloku)
  ošetření – pevné efekty β_j $(\sum_{j=1}^{r} \beta_j = 0)$ (vliv ošetření)

$$Y_{ij} = \mu + A_i + \beta_j + E_{ij}, \quad E_{ij} \sim \mathrm{N}(0, \sigma^2), \quad j = 1, \ldots, r, \quad i = 1, \ldots, k$$

předpokládá se **aditivní** vliv, symbolicky zapisovaný $A + B$

náhodné bloky

- zobecnění párových testů na r-tice
- **náhodný blok**
  - homogenní skupina r objektů
  - počet objektů ve skupině = počet ošetření (nebo jeho násobek)
  - ošetření se přiřadí uvnitř bloku **náhodně** (každému ošetření stejný počet objektů)
- bloky – náhodné efekty $A_i \sim \mathrm{N}(0, \sigma_A^2)$ (vliv bloku)
  ošetření – pevné efekty β_j $\left(\sum_{j=1}^{r} \beta_j = 0\right)$ (vliv ošetření)

$$Y_{ij} = \mu + A_i + \beta_j + E_{ij}, \quad E_{ij} \sim \mathrm{N}(0, \sigma^2), \quad j = 1, \ldots, r, \quad i = 1, \ldots, k$$

předpokládá se **aditivní** vliv, symbolicky zapisovaný $A + B$

náhodné bloky

- zobecnění párových testů na r-tice
- **náhodný blok**
  - homogenní skupina r objektů
  - počet objektů ve skupině = počet ošetření (nebo jeho násobek)
  - ošetření se přiřadí uvnitř bloku **náhodně** (každému ošetření stejný počet objektů)
- bloky – náhodné efekty $A_i \sim \mathrm{N}(0, \sigma_A^2)$ (vliv bloku)
  ošetření – pevné efekty β_j $(\sum_{j=1}^r \beta_j = 0)$ (vliv ošetření)

$$Y_{ij} = \mu + A_i + \beta_j + E_{ij}, \quad E_{ij} \sim \mathrm{N}(0, \sigma^2), \quad j = 1, \ldots, r, \quad i = 1, \ldots, k$$

předpokládá se **aditivní** vliv, symbolicky zapisovaný $A + B$

náhodné bloky

- zobecnění párových testů na r-tice
- **náhodný blok**
  - homogenní skupina r objektů
  - počet objektů ve skupině = počet ošetření (nebo jeho násobek)
  - ošetření se přiřadí uvnitř bloku **náhodně** (každému ošetření stejný počet objektů)
- bloky – náhodné efekty $A_i \sim \mathrm{N}(0, \sigma_A^2)$ (vliv bloku)
  ošetření – pevné efekty β_j $(\sum_{j=1}^{r} \beta_j = 0)$ (vliv ošetření)

$$Y_{ij} = \mu + A_i + \beta_j + E_{ij}, \quad E_{ij} \sim \mathrm{N}(0, \sigma^2), \quad j = 1, \ldots, r, \quad i = 1, \ldots, k$$

předpokládá se **aditivní** vliv, symbolicky zapisovaný $A + B$

náhodné bloky

- zobecnění párových testů na r-tice
- **náhodný blok**
  - homogenní skupina r objektů
  - počet objektů ve skupině = počet ošetření (nebo jeho násobek)
  - ošetření se přiřadí uvnitř bloku **náhodně** (každému ošetření stejný počet objektů)
- bloky – náhodné efekty $A_i \sim \mathrm{N}\left(0, \sigma_A^2\right)$ (vliv bloku)
  ošetření – pevné efekty β_j $\left(\sum_{j=1}^{r} \beta_j = 0\right)$ (vliv ošetření)

$$Y_{ij} = \mu + A_i + \beta_j + E_{ij}, \quad E_{ij} \sim \mathrm{N}\left(0, \sigma^2\right), \quad j = 1, \ldots, r, \quad i = 1, \ldots, k$$

předpokládá se **aditivní** vliv, symbolicky zapisovaný $A + B$

náhodné bloky

- testované hypotézy
  - $H_A : \sigma_A^2 = 0$ (nulová variabilita mezi bloky)
  - $H_B : \beta_1 = \ldots = \beta_r = 0$ (ošetření B nemá vliv)
- rozklad variability

$$S_T = S_A + S_B + S_e$$

- vliv dvou **faktorů**
  - A – náhodný: nastavuje příroda, při opakování pokusu budou úrovně jiné
  - B – pevný: nastavuje experimentátor, při opakování pokusu budou úrovně stejné

náhodné bloky

- testované hypotézy
  - $H_A : \sigma_A^2 = 0$ (nulová variabilita mezi bloky)
  - $H_B : \beta_1 = \ldots = \beta_r = 0$ (ošetření B nemá vliv)
- rozklad variability

$$S_T = S_A + S_B + S_e$$

- vliv dvou **faktorů**
  - A – náhodný: nastavuje příroda, při opakování pokusu budou úrovně jiné
  - B – pevný: nastavuje experimentátor, při opakování pokusu budou úrovně stejné

náhodné bloky

- testované hypotézy
  - $H_A : \sigma_A^2 = 0$ (nulová variabilita mezi bloky)
  - $H_B : \beta_1 = \ldots = \beta_r = 0$ (ošetření B nemá vliv)
- rozklad variability

$$S_T = S_A + S_B + S_e$$

- vliv dvou **faktorů**
  - A – náhodný: nastavuje příroda, při opakování pokusu budou úrovně jiné
  - B – pevný: nastavuje experimentátor, při opakování pokusu budou úrovně stejné

náhodné bloky

- testované hypotézy
  - $H_A : \sigma_A^2 = 0$ (nulová variabilita mezi bloky)
  - $H_B : \beta_1 = \ldots = \beta_r = 0$ (ošetření B nemá vliv)
- rozklad variability

$$S_T = S_A + S_B + S_e$$

- vliv dvou **faktorů**
  - A – náhodný: nastavuje příroda, při opakování pokusu budou úrovně jiné
  - B – pevný: nastavuje experimentátor, při opakování pokusu budou úrovně stejné

náhodné bloky

- testované hypotézy
  - $H_A : \sigma^2_A = 0$ (nulová variabilita mezi bloky)
  - $H_B : \beta_1 = \ldots = \beta_r = 0$ (ošetření B nemá vliv)
- rozklad variability

$$S_T = S_A + S_B + S_e$$

- vliv dvou **faktorů**
  - A – náhodný: nastavuje příroda, při opakování pokusu budou úrovně jiné
  - B – pevný: nastavuje experimentátor, při opakování pokusu budou úrovně stejné

náhodné bloky

- testované hypotézy
  - $H_A : \sigma_A^2 = 0$ (nulová variabilita mezi bloky)
  - $H_B : \beta_1 = \ldots = \beta_r = 0$ (ošetření B nemá vliv)
- rozklad variability

$$S_T = S_A + S_B + S_e$$

- vliv dvou **faktorů**
  - A – náhodný: nastavuje příroda, při opakování pokusu budou úrovně jiné
  - B – pevný: nastavuje experimentátor, při opakování pokusu budou úrovně stejné

náhodné bloky

- testované hypotézy
  - $H_A : \sigma_A^2 = 0$ (nulová variabilita mezi bloky)
  - $H_B : \beta_1 = \ldots = \beta_r = 0$ (ošetření B nemá vliv)
- rozklad variability

$$S_T = S_A + S_B + S_e$$

- vliv dvou **faktorů**
  - A – náhodný: nastavuje příroda, při opakování pokusu budou úrovně jiné
  - B – pevný: nastavuje experimentátor, při opakování pokusu budou úrovně stejné

příklad diety: váhové přírůstky za danou dobu

vrh	dieta				
	A	B	C	D	průměr
1	6,6	5,2	7,4	9,1	7,075
2	10,1	11,4	13,0	12,6	11,775
3	5,8	4,2	9,5	8,8	7,075
4	12,1	10,7	11,9	13,0	11,925
5	8,2	8,8	9,6	9,4	9,000
průměr	8,56	8,06	10,28	10,58	9,370

- $r = 4$ ošetření (pevné efekty, zvolili jsme je sami)
- $k = 5$ vrhů (náhodné efekty, zvolila je náhodně příroda)
- jsou patrné rozdíly mezi průměry pro jednotlivá ošetření i pro jednotlivé vrhy

příklad diety: váhové přírůstky za danou dobu

	dieta				
vrh	A	B	C	D	průměr
1	6,6	5,2	7,4	9,1	7,075
2	10,1	11,4	13,0	12,6	11,775
3	5,8	4,2	9,5	8,8	7,075
4	12,1	10,7	11,9	13,0	11,925
5	8,2	8,8	9,6	9,4	9,000
průměr	8,56	8,06	10,28	10,58	9,370

- $r = 4$ ošetření (pevné efekty, zvolili jsme je sami)
- $k = 5$ vrhů (náhodné efekty, zvolila je náhodně příroda)
- jsou patrné rozdíly mezi průměry pro jednotlivá ošetření i pro jednotlivé vrhy

příklad diety: váhové přírůstky za danou dobu

	dieta				
vrh	A	B	C	D	průměr
1	6,6	5,2	7,4	9,1	7,075
2	10,1	11,4	13,0	12,6	11,775
3	5,8	4,2	9,5	8,8	7,075
4	12,1	10,7	11,9	13,0	11,925
5	8,2	8,8	9,6	9,4	9,000
průměr	8,56	8,06	10,28	10,58	9,370

- $r = 4$ ošetření (pevné efekty, zvolili jsme je sami)
- $k = 5$ vrhů (náhodné efekty, zvolila je náhodně příroda)
- jsou patrné rozdíly mezi průměry pro jednotlivá ošetření i pro jednotlivé vrhy

příklad diety

tabulka ANOVA

variabilita	S	f	S/f	F	p
vrhy	91,932	4	22,983	(22,26)	(<0,0001)
dieta	23,332	3	7,774	7,53	0,0043
reziduální	12,388	12	1,032	-	-
celk.	127,642	19	-	-	-

[summary(aov(prirustek∼Error(Vrh)+Dieta,data=Mysi))]
nesprávně aplikované jednoduché třídění ANOVA:
kdybychom nevzali v úvahu závislost některých pozorování způsobenou náhodnými bloky (vrhy), dostali bychom:
$S_e = 91{,}932 + 12{,}388 = 104{,}320$, $f_e = 4 + 12 = 16$

$$F = \frac{23{,}332/3}{104{,}320/16} = 1{,}193, \qquad p = 0{,}344$$

Friedmanův test
(neparametrický test)

- model $Y_{ij} = \mu + A_i + \beta_j + E_{ij}$ (náhodný řádkový efekt) nebo $Y_{ij} = \mu + \alpha_i + \beta_j + E_{ij}$ (pevný řádkový efekt)
- E_{ij} nezávislé, spojité rozdělení
- $\mathrm{H}_0 : \beta_1 = \ldots = \beta_r$ (nezávisí na ošetření)
- urči pořadí v rámci každého bloku (řádku) R_{ij}
- za hypotézy je v každém řádku náhodná permutace čísel $1,\ldots,r$, jsou součty ve sloupcích (pro ošetření) podobné
- $$Q = \frac{12}{kr(r+1)} \sum_{j=1}^{r} \left(\sum_{i=1}^{k} R_{ij} \right)^2 - 3k(r+1)$$
- zamítat H_0 : pro $Q \geq \chi^2_{r-1}(\alpha)$

Friedmanův test
(neparametrický test)

- model $Y_{ij} = \mu + A_i + \beta_j + E_{ij}$ (náhodný řádkový efekt) nebo $Y_{ij} = \mu + \alpha_i + \beta_j + E_{ij}$ (pevný řádkový efekt)
- E_{ij} nezávislé, spojité rozdělení
- $\mathrm{H}_0 : \beta_1 = \ldots = \beta_r$ (nezávisí na ošetření)
- urči pořadí v rámci každého bloku (řádku) R_{ij}
- za hypotézy je v každém řádku náhodná permutace čísel $1, \ldots, r$, jsou součty ve sloupcích (pro ošetření) podobné
- $$Q = \frac{12}{kr(r+1)} \sum_{j=1}^{r} \left(\sum_{i=1}^{k} R_{ij} \right)^2 - 3k(r+1)$$
- zamítat H_0 : pro $Q \geq \chi^2_{r-1}(\alpha)$

Friedmanův test
(neparametrický test)

- model $Y_{ij} = \mu + A_i + \beta_j + E_{ij}$ (náhodný řádkový efekt) nebo $Y_{ij} = \mu + \alpha_i + \beta_j + E_{ij}$ (pevný řádkový efekt)
- E_{ij} nezávislé, spojité rozdělení
- $\mathrm{H}_0 : \beta_1 = \ldots = \beta_r$ (nezávisí na ošetření)
- urči pořadí v rámci každého bloku (řádku) R_{ij}
- za hypotézy je v každém řádku náhodná permutace čísel $1, \ldots, r$, jsou součty ve sloupcích (pro ošetření) podobné
- $$Q = \frac{12}{kr(r+1)} \sum_{j=1}^{r} \left(\sum_{i=1}^{k} R_{ij} \right)^2 - 3k(r+1)$$
- zamítat H_0 : pro $Q \geq \chi^2_{r-1}(\alpha)$

Friedmanův test
(neparametrický test)

- model $Y_{ij} = \mu + A_i + \beta_j + E_{ij}$ (náhodný řádkový efekt) nebo $Y_{ij} = \mu + \alpha_i + \beta_j + E_{ij}$ (pevný řádkový efekt)
- E_{ij} nezávislé, spojité rozdělení
- $\mathrm{H}_0 : \beta_1 = \ldots = \beta_r$ (nezávisí na ošetření)
- urči pořadí v rámci každého bloku (řádku) R_{ij}
- za hypotézy je v každém řádku náhodná permutace čísel $1,\ldots,r$, jsou součty ve sloupcích (pro ošetření) podobné
- $$Q = \frac{12}{kr(r+1)} \sum_{j=1}^{r} \left(\sum_{i=1}^{k} R_{ij} \right)^2 - 3k(r+1)$$
- zamítat H_0 : pro $Q \geq \chi^2_{r-1}(\alpha)$

Friedmanův test

(neparametrický test)

- model $Y_{ij} = \mu + A_i + \beta_j + E_{ij}$ (náhodný řádkový efekt) nebo $Y_{ij} = \mu + \alpha_i + \beta_j + E_{ij}$ (pevný řádkový efekt)
- E_{ij} nezávislé, spojité rozdělení
- $\mathrm{H}_0 : \beta_1 = \ldots = \beta_r$ (nezávisí na ošetření)
- urči pořadí v rámci každého bloku (řádku) R_{ij}
- za hypotézy je v každém řádku náhodná permutace čísel $1, \ldots, r$, jsou součty ve sloupcích (pro ošetření) podobné
- $$Q = \frac{12}{kr(r+1)} \sum_{j=1}^{r} \left(\sum_{i=1}^{k} R_{ij} \right)^2 - 3k(r+1)$$
- zamítat H_0 : pro $Q \geq \chi^2_{r-1}(\alpha)$

Friedmanův test
(neparametrický test)

- model $Y_{ij} = \mu + A_i + \beta_j + E_{ij}$ (náhodný řádkový efekt) nebo $Y_{ij} = \mu + \alpha_i + \beta_j + E_{ij}$ (pevný řádkový efekt)
- E_{ij} nezávislé, spojité rozdělení
- $\mathrm{H}_0 : \beta_1 = \ldots = \beta_r$ (nezávisí na ošetření)
- urči pořadí v rámci každého bloku (řádku) R_{ij}
- za hypotézy je v každém řádku náhodná permutace čísel $1, \ldots, r$, jsou součty ve sloupcích (pro ošetření) podobné
- $$Q = \frac{12}{kr(r+1)} \sum_{j=1}^{r} \left(\sum_{i=1}^{k} R_{ij} \right)^2 - 3k(r+1)$$
- zamítat H_0 : pro $Q \geq \chi^2_{r-1}(\alpha)$

Friedmanův test
(neparametrický test)

- model $Y_{ij} = \mu + A_i + \beta_j + E_{ij}$ (náhodný řádkový efekt) nebo $Y_{ij} = \mu + \alpha_i + \beta_j + E_{ij}$ (pevný řádkový efekt)
- E_{ij} nezávislé, spojité rozdělení
- $\mathrm{H}_0 : \beta_1 = \ldots = \beta_r$ (nezávisí na ošetření)
- urči pořadí v rámci každého bloku (řádku) R_{ij}
- za hypotézy je v každém řádku náhodná permutace čísel $1, \ldots, r$, jsou součty ve sloupcích (pro ošetření) podobné
- $$Q = \frac{12}{kr(r+1)} \sum_{j=1}^{r} \left(\sum_{i=1}^{k} R_{ij} \right)^2 - 3k(r+1)$$
- zamítat H_0 : pro $Q \geq \chi^2_{r-1}(\alpha)$

příklad diety

[friedman.test(prirustek~Dieta|Vrh,data=Mysi)]

vrh	dieta A	B	C	D	prům.
1	6,6	5,2	7,4	9,1	7,075
2	10,1	11,4	13,0	12,6	11,775
3	5,8	4,2	9,5	8,8	7,075
4	12,1	10,7	11,9	13,0	11,925
5	8,2	8,8	9,6	9,4	9,000
prům.	8,56	8,06	10,28	10,58	9,370

vrh	dieta A	B	C	D	
1	2	1	3	4	
2	1	2	4	3	
3	2	1	4	3	
4	3	1	2	4	
5	1	2	4	3	
součet	9	7	17	17	

$$k = 5$$

$$r = 4$$

$$Q = \frac{12}{5 \cdot 4 \cdot 5}\left(9^2 + 7^2 + 17^2 + 17^2\right) - 3 \cdot 5 \cdot 6 = 9{,}96$$

$$Q > \chi^2_3(0{,}05) = 7{,}8147$$

$$p = 0{,}0189$$

dvojné třídění s interakcemi

- vliv dvou faktorů nemusí být aditivní

$$Y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_{ij} + E_{ijt}$$
$$E_{ijt} \sim \mathrm{N}\left(0, \sigma^2\right)$$

- symbolicky $A + B + AB$
- $\sum_i \alpha_i = 0$
  efekty faktoru A odpovídající jeho k úrovním
- $\sum_j \beta_j = 0$
  efekty faktoru B odpovídající jeho r úrovním
- $\sum_i \gamma_{ij} = 0, \quad \sum_j \gamma_{ij} = 0$
  interakce vyjadřují neaditivitu obou faktorů
  (vliv A závisí na úrovni B, vliv B závisí na úrovni A)

dvojné třídění s interakcemi

- vliv dvou faktorů nemusí být aditivní

$$Y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_{ij} + E_{ijt}$$
$$E_{ijt} \sim \mathrm{N}(0, \sigma^2)$$

- symbolicky $A + B + AB$
- $\sum_i \alpha_i = 0$
  efekty faktoru A odpovídající jeho k úrovním
- $\sum_j \beta_j = 0$
  efekty faktoru B odpovídající jeho r úrovním
- $\sum_i \gamma_{ij} = 0, \quad \sum_j \gamma_{ij} = 0$
  interakce vyjadřují neaditivitu obou faktorů
  (vliv A závisí na úrovni B, vliv B závisí na úrovni A)

dvojné třídění s interakcemi

- vliv dvou faktorů nemusí být aditivní

$$Y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_{ij} + E_{ijt}$$
$$E_{ijt} \sim \mathrm{N}\left(0, \sigma^2\right)$$

- symbolicky $A + B + AB$
- $\sum_i \alpha_i = 0$
  efekty faktoru A odpovídající jeho k úrovním
- $\sum_j \beta_j = 0$
  efekty faktoru B odpovídající jeho r úrovním
- $\sum_i \gamma_{ij} = 0, \quad \sum_j \gamma_{ij} = 0$
  interakce vyjadřují neaditivitu obou faktorů
  (vliv A závisí na úrovni B, vliv B závisí na úrovni A)

dvojné třídění s interakcemi

- vliv dvou faktorů nemusí být aditivní

$$Y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_{ij} + E_{ijt}$$
$$E_{ijt} \sim \mathrm{N}\left(0, \sigma^2\right)$$

- symbolicky $A + B + AB$
- $\sum_i \alpha_i = 0$
  efekty faktoru A odpovídající jeho k úrovním
- $\sum_j \beta_j = 0$
  efekty faktoru B odpovídající jeho r úrovním
- $\sum_i \gamma_{ij} = 0, \quad \sum_j \gamma_{ij} = 0$
  interakce vyjadřují neaditivitu obou faktorů
  (vliv A závisí na úrovni B, vliv B závisí na úrovni A)

dvojné třídění s interakcemi

- vliv dvou faktorů nemusí být aditivní

$$Y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_{ij} + E_{ijt}$$
$$E_{ijt} \sim \mathrm{N}\big(0, \sigma^2\big)$$

- symbolicky $A + B + AB$
- $\sum_i \alpha_i = 0$
  efekty faktoru A odpovídající jeho k úrovním
- $\sum_j \beta_j = 0$
  efekty faktoru B odpovídající jeho r úrovním
- $\sum_i \gamma_{ij} = 0, \quad \sum_j \gamma_{ij} = 0$
  interakce vyjadřují neaditivitu obou faktorů
  (vliv A závisí na úrovni B, vliv B závisí na úrovni A)

testy ve dvojném třídění

- $\mathrm{H}_{AB} : \gamma_{ij} = 0$ (aditivita obou faktorů)
  vliv úrovně faktoru A je stejný při všech úrovních faktoru B
  vliv úrovně faktoru B je stejný při všech úrovních faktoru A
- $\mathrm{H}_A : \alpha_i = 0$ (faktor A nemá vliv)
- $\mathrm{H}_B : \beta_j = 0$ (faktor B nemá vliv)
- pokud zamítneme H_{AB}, nemá smysl testovat $\mathrm{H}_A, \mathrm{H}_B$, neboť prostřednictvím interakcí oba faktory vliv mají
- pak je lépe přejít k modelu jednoduchého třídění s kombinovanými úrovněmi

testy ve dvojném třídění

- $\mathrm{H}_{AB} : \gamma_{ij} = 0$ (aditivita obou faktorů)
  vliv úrovně faktoru A je stejný při všech úrovních faktoru B
  vliv úrovně faktoru B je stejný při všech úrovních faktoru A
- $\mathrm{H}_{A} : \alpha_{i} = 0$ (faktor A nemá vliv)
- $\mathrm{H}_{B} : \beta_{j} = 0$ (faktor B nemá vliv)
- pokud zamítneme H_{AB}, nemá smysl testovat $\mathrm{H}_{A}, \mathrm{H}_{B}$, neboť prostřednictvím interakcí oba faktory vliv mají
- pak je lépe přejít k modelu jednoduchého třídění s kombinovanými úrovněmi

testy ve dvojném třídění

- $H_{AB} : \gamma_{ij} = 0$ (aditivita obou faktorů)
  vliv úrovně faktoru A je stejný při všech úrovních faktoru B
  vliv úrovně faktoru B je stejný při všech úrovních faktoru A
- $H_A : \alpha_i = 0$ (faktor A nemá vliv)
- $H_B : \beta_j = 0$ (faktor B nemá vliv)
- pokud zamítneme H_{AB}, nemá smysl testovat H_A, H_B, neboť prostřednictvím interakcí oba faktory vliv mají
- pak je lépe přejít k modelu jednoduchého třídění s kombinovanými úrovněmi

testy ve dvojném třídění

- $H_{AB} : \gamma_{ij} = 0$ (aditivita obou faktorů)
  vliv úrovně faktoru A je stejný při všech úrovních faktoru B
  vliv úrovně faktoru B je stejný při všech úrovních faktoru A
- $H_A : \alpha_i = 0$ (faktor A nemá vliv)
- $H_B : \beta_j = 0$ (faktor B nemá vliv)
- pokud zamítneme H_{AB}, nemá smysl testovat H_A, H_B, neboť prostřednictvím interakcí oba faktory vliv mají
- pak je lépe přejít k modelu jednoduchého třídění s kombinovanými úrovněmi

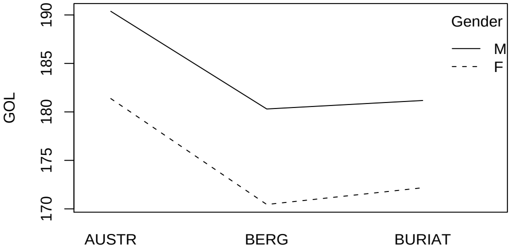
testy ve dvojném třídění

- $\mathrm{H}_{AB} : \gamma_{ij} = 0$ (aditivita obou faktorů)
  vliv úrovně faktoru A je stejný při všech úrovních faktoru B
  vliv úrovně faktoru B je stejný při všech úrovních faktoru A
- $\mathrm{H}_{A} : \alpha_{i} = 0$ (faktor A nemá vliv)
- $\mathrm{H}_{B} : \beta_{j} = 0$ (faktor B nemá vliv)
- pokud zamítneme H_{AB}, nemá smysl testovat $\mathrm{H}_{A}, \mathrm{H}_{B}$, neboť prostřednictvím interakcí oba faktory vliv mají
- pak je lépe přejít k modelu jednoduchého třídění
  s kombinovanými úrovněmi

příklad Howells

- lebky exhumované na třech místech (A)
- lebky jsou rozlišovány podle pohlaví (B)
- měříme největší délku mozkovny GOL

[anova(lm(gol∼Gender*Popul))] nebo
[anova(lm(gol∼Gender+Popul+Gender:Popul))]

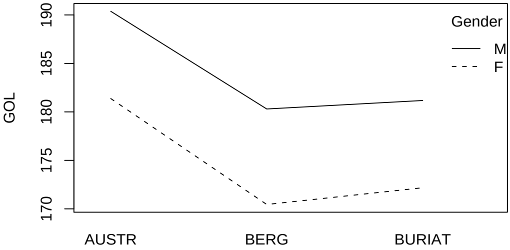


$$p_{AB} = 0{,}8872$$

příklad Howells

- lebky exhumované na třech místech (A)
- lebky jsou rozlišovány podle pohlaví (B)
- měříme největší délku mozkovny GOL

[anova(lm(gol~Gender*Popul))] nebo [anova(lm(gol~Gender+Popul+Gender:Popul))]

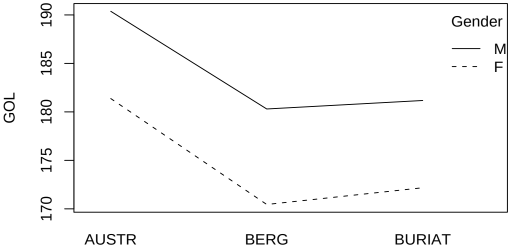


$$p_{AB} = 0{,}8872$$

příklad Howells

- lebky exhumované na třech místech (A)
- lebky jsou rozlišovány podle pohlaví (B)
- měříme největší délku mozkovny GOL

[anova(lm(gol~Gender*Popul))] nebo
[anova(lm(gol~Gender+Popul+Gender:Popul))]



$$p_{AB} = 0{,}8872$$

příklad Howells (GOL)

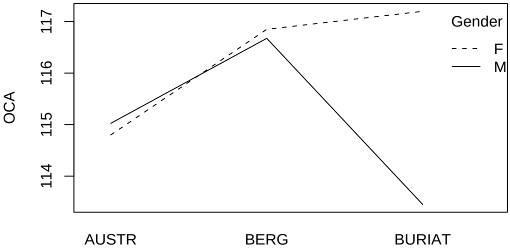
pohlaví	místo	n_{ij}	$\bar{y}_{ij}$	s_{ij}
M	Berg	40	180,300	7,293
F	Berg	40	170,450	6,641
M	Austrálie	40	190,375	5,555
F	Austrálie	40	181,375	6,632
M	Sibiř	40	181,175	6,468
F	Sibiř	40	172,175	5,228

var.	S	f	S/f	F	p
místa	5242,1	2	2621,1	65,2	$<$0,0001
pohl.	5170,8	1	5170,8	128,6	$<$0,0001
inter.	9,6	2	4,8	0,1	0,8872
rezid.	9410,6	234	40,2		
celk.	19833,2	239			

příklad Howells

- lebky exhumované na třech místech (A)
- lebky jsou rozlišovány podle pohlaví (B)
- měříme týlní úhel OCA

[anova(lm(oca~Gender*Popul))] nebo
[anova(lm(oca~Gender+Popul+Gender:Popul))]

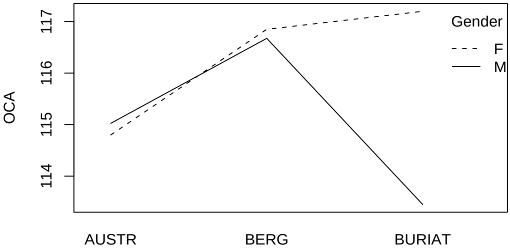


$p_{AB} = 0{,}0222$

příklad Howells

- lebky exhumované na třech místech (A)
- lebky jsou rozlišovány podle pohlaví (B)
- měříme týlní úhel OCA

[anova(lm(oca~Gender*Popul))] nebo
[anova(lm(oca~Gender+Popul+Gender:Popul))]

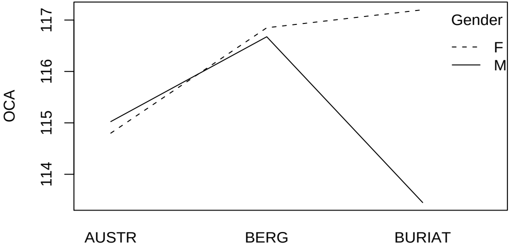


$$p_{AB} = 0{,}0222$$

příklad Howells

- lebky exhumované na třech místech (A)
- lebky jsou rozlišovány podle pohlaví (B)
- měříme týlní úhel OCA

[anova(lm(oca∼Gender*Popul))] nebo
[anova(lm(oca∼Gender+Popul+Gender:Popul))]



$$p_{AB} = 0{,}0222$$

příklad Howells (OCA)

pohlaví	místo	n_{ij}	$\bar{y}_{ij}$	s_{ij}
M	Berg	40	116,675	5,567
F	Berg	40	116,850	5,682
M	Austrálie	40	115,025	4,382
F	Austrálie	40	114,800	4,286
M	Sibiř	40	113,450	4,782
F	Sibiř	40	117,200	4,973

var.	S	f	S/f	F	p
místa	150,908	2	75,454	3,05	0,0493
pohl.	91,267	1	91,267	3,69	0,0560
inter.	191,608	2	95,804	3,87	0,0222
rezid.	5789,550	234	24,742		
celk.	6223,333	239			

porovnání populačních měr polohy

rozdělení	normální	spojité
populační parametr (o čem je hypotéza)	populační průměr	populační medián (distribuční funkce)
jeden výběr	jednovýběrový t-test	jednovýběrový Wilcoxon
výběr dvojic	párový t-test	znaménkový Wilcoxon
dva nezávislé výběry	dvouvýběrový t-test	Mann-Whitney (Kolmogorov-Smirnov)
k nezávislých výběrů	analýza rozptylu jedn. třídění	Kruskal-Wallis
výběr r-tic	analýza rozptylu náhodné bloky	Friedman