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 21. dubna 2009)

motivační 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
- kdyby byly jen dvě diety ($r = 2$), použili bychom párový test

motivační 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
- kdyby byly jen dvě diety ($r = 2$), použili bychom párový test

motivační 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
- kdyby byly jen dvě diety ($r = 2$), použili bychom párový test

motivační příklad diety: váhové přírůstky za danou dobu

vrh	dieta				průměr
	A	B	C	D	
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
- kdyby byly jen dvě diety ($r = 2$), použili bychom párový test

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)

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

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 $\left(\sum_{j=1}^{r} \beta_j = 0\right)$ (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 $\left(\sum_{j=1}^{r} \beta_j = 0\right)$ (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 $\left(\sum_{j=1}^{r} \beta_j = 0\right)$ (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 $A_i \sim \mathrm{N}\left(0, \sigma_A^2\right)$ (vliv bloku)
  ošetření – pevné efekty β_j $\left(\sum_{j=1}^{r} \beta_j = 0\right)$ (vliv ošetření)

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

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

náhodné bloky

- testované hypotézy
  - $H_B : \beta_1 = \ldots = \beta_r = 0$ (ošetření B nemá vliv)
  - případně $H_A : \sigma_A^2 = 0$ (nulová variabilita mezi bloky)
- 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
  - rozhodování zda A je pevný nebo náhodný efekt závisí na cíli výzkumu, na interpretaci

náhodné bloky

- testované hypotézy
  - $H_B : \beta_1 = \ldots = \beta_r = 0$ (ošetření B nemá vliv)
  - případně $H_A : \sigma_A^2 = 0$ (nulová variabilita mezi bloky)
- 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
  - rozhodování zda A je pevný nebo náhodný efekt závisí na cíli výzkumu, na interpretaci

náhodné bloky

- testované hypotézy
  - $H_B : \beta_1 = \ldots = \beta_r = 0$ (ošetření B nemá vliv)
  - případně $H_A : \sigma_A^2 = 0$ (nulová variabilita mezi bloky)
- 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
  - rozhodování zda A je pevný nebo náhodný efekt závisí na cíli výzkumu, na interpretaci

náhodné bloky

- testované hypotézy
  - $H_B : \beta_1 = \ldots = \beta_r = 0$ (ošetření B nemá vliv)
  - případně $H_A : \sigma_A^2 = 0$ (nulová variabilita mezi bloky)
- 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
  - rozhodování zda A je pevný nebo náhodný efekt závisí na cíli výzkumu, na interpretaci

náhodné bloky

- testované hypotézy
  - $H_B : \beta_1 = \ldots = \beta_r = 0$ (ošetření B nemá vliv)
  - případně $H_A : \sigma_A^2 = 0$ (nulová variabilita mezi bloky)
- 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
  - rozhodování zda A je pevný nebo náhodný efekt závisí na cíli výzkumu, na interpretaci

náhodné bloky

- testované hypotézy
  - $H_B : \beta_1 = \ldots = \beta_r = 0$ (ošetření B nemá vliv)
  - případně $H_A : \sigma_A^2 = 0$ (nulová variabilita mezi bloky)
- 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
  - rozhodování zda A je pevný nebo náhodný efekt závisí na cíli výzkumu, na interpretaci

náhodné bloky

- testované hypotézy
  - $H_B : \beta_1 = \ldots = \beta_r = 0$ (ošetření B nemá vliv)
  - případně $H_A : \sigma_A^2 = 0$ (nulová variabilita mezi bloky)
- 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
  - rozhodování zda A je pevný nebo náhodný efekt závisí na cíli výzkumu, na interpretaci

náhodné bloky

- testované hypotézy
  - $H_B : \beta_1 = \ldots = \beta_r = 0$ (ošetření B nemá vliv)
  - případně $H_A : \sigma_A^2 = 0$ (nulová variabilita mezi bloky)
- 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
  - rozhodování zda A je pevný nebo náhodný efekt závisí na cíli výzkumu, na interpretaci

příklad diety

- tabulka ANOVA

variabilita	S	f	S/f	F	p
vrhy	91,932	4	22,983	(22,26)	(<0,0001)
dieta	23,322	3	7,774	7,53	0,0043
reziduální	12,388	12	1,032	-	-
celk.	127,642	19	-	-	-

- na 5% hladině jsme prokázali rozdíl mezi dietami ($p = 0{,}4$ %)
- variabilita mezi vrhy je také průkazná ($p < 0{,}1$ %)
- [summary(aov(prirustek~Error(Vrh)+Dieta,data=Mysi))]
- pro takto jednoduchý model vyjde tabulka stejně i když považujeme faktor A za pevný (nenáhodný); porovnáváme pak konkrétních pět vrhů, vrhy nechápeme jako vzorek všech možných vrhů

příklad diety

- tabulka ANOVA

variabilita	S	f	S/f	F	p
vrhy	91,932	4	22,983	(22,26)	(<0,0001)
dieta	23,322	3	7,774	7,53	0,0043
reziduální	12,388	12	1,032	-	-
celk.	127,642	19	-	-	-

- na 5% hladině jsme prokázali rozdíl mezi dietami ($p = 0{,}4$ %)
- variabilita mezi vrhy je také průkazná ($p < 0{,}1$ %)
- [summary(aov(prirustek~Error(Vrh)+Dieta,data=Mysi))]
- pro takto jednoduchý model vyjde tabulka stejně i když považujeme faktor A za pevný (nenáhodný); porovnáváme pak konkrétních pět vrhů, vrhy nechápeme jako vzorek všech možných vrhů

příklad diety

- tabulka ANOVA

variabilita	S	f	S/f	F	p
vrhy	91,932	4	22,983	(22,26)	(<0,0001)
dieta	23,322	3	7,774	7,53	0,0043
reziduální	12,388	12	1,032	-	-
celk.	127,642	19	-	-	-

- na 5% hladině jsme prokázali rozdíl mezi dietami ($p = 0{,}4$ %)
- variabilita mezi vrhy je také průkazná ($p < 0{,}1$ %)
- [summary(aov(prirustek~Error(Vrh)+Dieta,data=Mysi))]
- pro takto jednoduchý model vyjde tabulka stejně i když považujeme faktor A za pevný (nenáhodný); porovnáváme pak konkrétních pět vrhů, vrhy nechápeme jako vzorek všech možných vrhů

příklad diety

- tabulka ANOVA

variabilita	S	f	S/f	F	p
vrhy	91,932	4	22,983	(22,26)	(<0,0001)
dieta	23,322	3	7,774	7,53	0,0043
reziduální	12,388	12	1,032	-	-
celk.	127,642	19	-	-	-

- na 5% hladině jsme prokázali rozdíl mezi dietami ($p = 0{,}4$ %)
- variabilita mezi vrhy je také průkazná ($p < 0{,}1$ %)
- [summary(aov(prirustek~Error(Vrh)+Dieta,data=Mysi))]
- pro takto jednoduchý model vyjde tabulka stejně i když považujeme faktor A za pevný (nenáhodný); porovnáváme pak konkrétních pět vrhů, vrhy nechápeme jako vzorek všech možných vrhů

příklad diety

- tabulka ANOVA

variabilita	S	f	S/f	F	p
vrhy	91,932	4	22,983	(22,26)	(<0,0001)
dieta	23,322	3	7,774	7,53	0,0043
reziduální	12,388	12	1,032	-	-
celk.	127,642	19	-	-	-

- na 5% hladině jsme prokázali rozdíl mezi dietami ($p = 0{,}4$ %)
- variabilita mezi vrhy je také průkazná ($p < 0{,}1$ %)
- [summary(aov(prirustek~Error(Vrh)+Dieta,data=Mysi))]
- pro takto jednoduchý model vyjde tabulka stejně i když považujeme faktor A za pevný (nenáhodný); porovnáváme pak konkrétních pět vrhů, vrhy nechápeme jako vzorek všech možných vrhů

příklad diety

- kdybychom **nesprávně** nevzali v úvahu závislost některých pozorování způsobenou náhodnými bloky (vrhy), dostali bychom model ANOVA jednoduchého třídění

variabilita	S	f	S/f	F	p
dieta	23,332	3	7,774	1,193	0,344
reziduální	104,320	16	6,520	-	-
celk.	127,642	19	-	-	-

- [summary(aov(prirustek~Dieta,data=Mysi))]
- porovnání se správnou tabulkou analýzy rozptylu

$$S_e = 91{,}932 + 12{,}388 = 104{,}320, \qquad f_e = 4 + 12 = 16$$

příklad diety

- kdybychom **nesprávně** nevzali v úvahu závislost některých pozorování způsobenou náhodnými bloky (vrhy), dostali bychom model ANOVA jednoduchého třídění

variabilita	S	f	S/f	F	p
dieta	23,332	3	7,774	1,193	0,344
reziduální	104,320	16	6,520	-	-
celk.	127,642	19	-	-	-

- [summary(aov(prirustek~Dieta,data=Mysi))]
- porovnání se správnou tabulkou analýzy rozptylu

$$S_e = 91{,}932 + 12{,}388 = 104{,}320, \qquad f_e = 4 + 12 = 16$$

příklad diety

- kdybychom **nesprávně** nevzali v úvahu závislost některých pozorování způsobenou náhodnými bloky (vrhy), dostali bychom model ANOVA jednoduchého třídění

variabilita	S	f	S/f	F	p
dieta	23,332	3	7,774	1,193	0,344
reziduální	104,320	16	6,520	-	-
celk.	127,642	19	-	-	-

- [summary(aov(prirustek∼Dieta,data=Mysi))]
- porovnání se správnou tabulkou analýzy rozptylu

$$S_e = 91{,}932 + 12{,}388 = 104{,}320, \quad f_e = 4 + 12 = 16$$

Friedmanův test, zobecnění znaménkového testu

(neparametrický test, bez předpokladu normality)

- 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 (nemusí být normální)
- $H_0 : \beta_1 = \ldots = \beta_r$ (nezávisí na ošetření)
- urči pořadí v rámci každého bloku (řádku) R_{ij}
- za hypotézy je v každém řádku náhodná permutace čísel $1,\ldots,r$, součty ve sloupcích (pro ošetření) jsou 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, zobecnění znaménkového testu

(neparametrický test, bez předpokladu normality)

- 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 (nemusí být normální)
- $H_0 : \beta_1 = \ldots = \beta_r$ (nezávisí na ošetření)
- urči pořadí v rámci každého bloku (řádku) R_{ij}
- za hypotézy je v každém řádku náhodná permutace čísel $1, \ldots, r$, součty ve sloupcích (pro ošetření) jsou 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, zobecnění znaménkového testu

(neparametrický test, bez předpokladu normality)

- 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 (nemusí být normální)
- $H_0 : \beta_1 = \ldots = \beta_r$ (nezávisí na ošetření)
- urči pořadí v rámci každého bloku (řádku) R_{ij}
- za hypotézy je v každém řádku náhodná permutace čísel $1, \ldots, r$, součty ve sloupcích (pro ošetření) jsou 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, zobecnění znaménkového testu

(neparametrický test, bez předpokladu normality)

- 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 (nemusí být normální)
- $H_0 : \beta_1 = \ldots = \beta_r$ (nezávisí na ošetření)
- urči pořadí v rámci každého bloku (řádku) R_{ij}
- za hypotézy je v každém řádku náhodná permutace čísel $1,\ldots,r$, součty ve sloupcích (pro ošetření) jsou 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, zobecnění znaménkového testu

(neparametrický test, bez předpokladu normality)

- 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 (nemusí být normální)
- $H_0 : \beta_1 = \ldots = \beta_r$ (nezávisí na ošetření)
- urči pořadí v rámci každého bloku (řádku) R_{ij}
- za hypotézy je v každém řádku náhodná permutace čísel $1, \ldots, r$, součty ve sloupcích (pro ošetření) jsou 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, zobecnění znaménkového testu
(neparametrický test, bez předpokladu normality)

- 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 (nemusí být normální)
- $H_0 : \beta_1 = \ldots = \beta_r$ (nezávisí na ošetření)
- urči pořadí v rámci každého bloku (řádku) R_{ij}
- za hypotézy je v každém řádku náhodná permutace čísel $1,\ldots,r$, součty ve sloupcích (pro ošetření) jsou 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, zobecnění znaménkového testu
(neparametrický test, bez předpokladu normality)

- 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 (nemusí být normální)
- $\mathrm{H}_0 : \beta_1 = \ldots = \beta_r$ (nezávisí na ošetření)
- urči pořadí v rámci každého bloku (řádku) R_{ij}
- za hypotézy je v každém řádku náhodná permutace čísel $1,\ldots,r$, součty ve sloupcích (pro ošetření) jsou 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	dieta B	dieta C	dieta 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	dieta B	dieta C	dieta 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	

$$k = 5$$

$$r = 4$$

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

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

$$p = 0{,}0189$$

dvojné třídění s interakcemi

opět **normální rozdělení**

- vliv dvou faktorů nemusí být aditivní ($1 \leq t \leq T$)

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

- 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), pak
  dvojné třídění bez interakcí (s opakováním pro $T > 1$)

dvojné třídění s interakcemi
opět **normální rozdělení**

- vliv dvou faktorů nemusí být aditivní ($1 \leq t \leq T$)

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

- 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), pak
  dvojné třídění bez interakcí (s opakováním pro $T > 1$)

dvojné třídění s interakcemi

opět **normální rozdělení**

- vliv dvou faktorů nemusí být aditivní ($1 \le t \le T$)

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

- 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), pak
  dvojné třídění bez interakcí (s opakováním pro $T > 1$)

dvojné třídění s interakcemi

opět **normální rozdělení**

- vliv dvou faktorů nemusí být aditivní ($1 \le t \le T$)

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

- 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), pak
  dvojné třídění bez interakcí (s opakováním pro $T > 1$)

dvojné třídění s interakcemi

opět **normální rozdělení**

- vliv dvou faktorů nemusí být aditivní ($1 \leq t \leq T$)

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

- 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), pak
  dvojné třídění bez interakcí (s opakováním pro $T > 1$)

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
- v takovém případě 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
- v takovém případě 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
- v takovém případě 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
- v takovém případě je lépe přejít k modelu jednoduchého třídění s kombinovanými úrovněmi

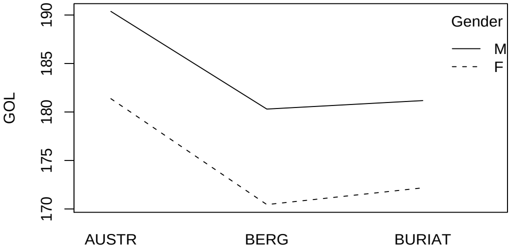
testy ve dvojném třídění

- $H_{AB} : \gamma_{ij} = 0$ (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
- v takovém případě 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))]

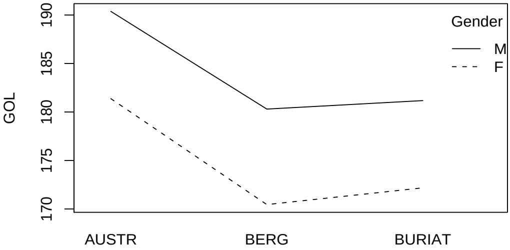


$$p_{AB} = 0{,}8872$$

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))]

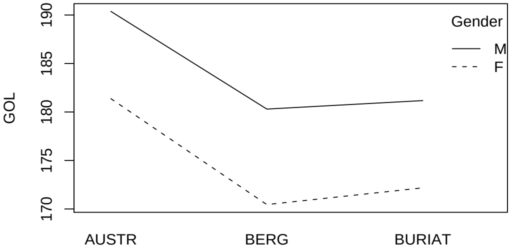


$$p_{AB} = 0{,}8872$$

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)

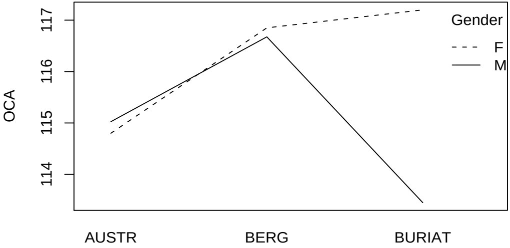
pohlaví	místo	n_{ij}	$\bar{y}_{ij}$	s_{ij}
M	Berg	40	180,300	7,293
F	Berg	40	170,450	6,641
M	Austrálie	40	190,375	5,555
F	Austrálie	40	181,375	6,632
M	Sibiř	40	181,175	6,468
F	Sibiř	40	172,175	5,228

variabilita	S	f	S/f	F	p
místa	5242,1	2	2621,1	65,2	$<0{,}0001$
pohlaví	5170,8	1	5170,8	128,6	$<0{,}0001$
interakce	9,6	2	4,8	0,1	0,8872
reziduální	9410,6	234	40,2		
celková	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))]

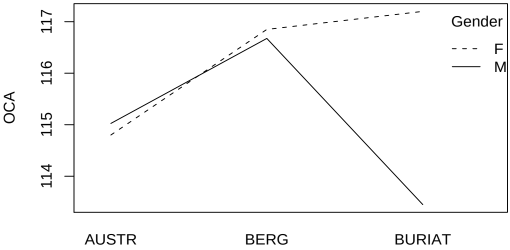


$$p_{AB} = 0{,}0222$$

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))]

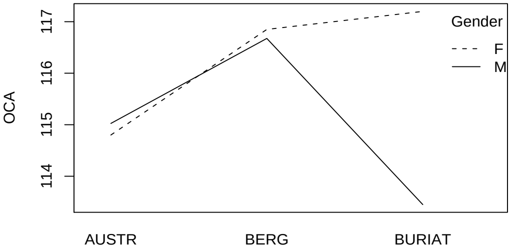


$$p_{AB} = 0{,}0222$$

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

variabilita	S	f	S/f	F	p
místa	150,908	2	75,454	3,05	0,0493
pohlaví	91,267	1	91,267	3,69	0,0560
interakce	191,608	2	95,804	3,87	0,0222
reziduální	5789,550	234	24,742		
celková	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

vyšetřování závislosti

nezávisle proměnná(é)	závisle proměnná	
	spojitá	**nominální**
spojitá	regrese korelace	(*logistická regrese*)
nominální	analýza rozptylu	kontingenční tabulky

příklady:

- hmotnost na výšce
- rakovina plic na počtu vykouřených cigaret
- hmotnost obilky na živném roztoku
- barva očí a barva vlasů

vyšetřování závislosti

nezávisle proměnná(é)	závisle proměnná	
	spojitá	**nominální**
spojitá	regrese korelace	(*logistická regrese*)
nominální	analýza rozptylu	kontingenční tabulky

příklady:

- hmotnost na výšce
- rakovina plic na počtu vykouřených cigaret
- hmotnost obilky na živném roztoku
- barva očí a barva vlasů

vyšetřování závislosti

nezávisle proměnná(é)	závisle proměnná	
	spojitá	**nominální**
spojitá	regrese korelace	(*logistická regrese*)
nominální	analýza rozptylu	kontingenční tabulky

příklady:

- hmotnost na výšce
- rakovina plic na počtu vykouřených cigaret
- hmotnost obilky na živném roztoku
- barva očí a barva vlasů

vyšetřování závislosti

nezávisle proměnná(é)	závisle proměnná	
	spojitá	**nominální**
spojitá	regrese korelace	(*logistická regrese*)
nominální	analýza rozptylu	kontingenční tabulky

příklady:

- hmotnost na výšce
- rakovina plic na počtu vykouřených cigaret
- hmotnost obilky na živném roztoku
- barva očí a barva vlasů

korelace a regrese

[correlation, regression]

- **korelace** (dvojice náhodných veličin)
  - měří **sílu** (těsnost) **vzájemné** závislosti **spojitých** veličin
  - lze použít k **prokazování** existence **vzájemné** závislosti X, Y
  - k **porovnávání síly** (těsnosti) závislosti v několika populacích
  - **symetrická** vlastnost veličin X a Y
- **regrese** (náhodná veličina na nenáhodné veličině)
  - udává **jak** závisí střední hodnota **spojité** veličiny Y na nezávisle proměnné (proměnných) x
  - **nesymetrická** vlastnost (závislost Y na $x \neq$ závislost X na y)
  - lze použít k **prokazování** existence závislosti **závisle** proměnné Y na **nezávisle** proměnné x
  - umožňuje **předpovídat** stř. hodnotu Y pro zvolenou hodnotu x

korelace a regrese

[correlation, regression]

- **korelace** (dvojice náhodných veličin)
  - měří **sílu** (těsnost) **vzájemné** závislosti **spojitých** veličin
  - lze použít k **prokazování** existence **vzájemné** závislosti X, Y
  - k **porovnávání síly** (těsnosti) závislosti v několika populacích
  - **symetrická** vlastnost veličin X a Y
- **regrese** (náhodná veličina na nenáhodné veličině)
  - udává **jak** závisí střední hodnota **spojité** veličiny Y na nezávisle proměnné (proměnných) x
  - **nesymetrická** vlastnost (závislost Y na $x \neq$ závislost X na y)
  - lze použít k **prokazování** existence závislosti **závisle** proměnné Y na **nezávisle** proměnné x
  - umožňuje **předpovídat** stř. hodnotu Y pro zvolenou hodnotu x

korelace a regrese

[correlation, regression]

- **korelace** (dvojice náhodných veličin)
  - měří **sílu** (těsnost) **vzájemné** závislosti **spojitých** veličin
  - lze použít k **prokazování** existence **vzájemné** závislosti X, Y
  - k **porovnávání síly** (těsnosti) závislosti v několika populacích
  - **symetrická** vlastnost veličin X a Y
- **regrese** (náhodná veličina na nenáhodné veličině)
  - udává **jak** závisí střední hodnota **spojité** veličiny Y na nezávisle proměnné (proměnných) x
  - **nesymetrická** vlastnost (závislost Y na $x \neq$ závislost X na y)
  - lze použít k **prokazování** existence závislosti **závisle** proměnné Y na **nezávisle** proměnné x
  - umožňuje **předpovídat** stř. hodnotu Y pro zvolenou hodnotu x

korelace a regrese
[correlation, regression]

- **korelace** (dvojice náhodných veličin)
  - měří **sílu** (těsnost) **vzájemné** závislosti **spojitých** veličin
  - lze použít k **prokazování** existence **vzájemné** závislosti X, Y
  - k **porovnávání síly** (těsnosti) závislosti v několika populacích
  - **symetrická** vlastnost veličin X a Y
- **regrese** (náhodná veličina na nenáhodné veličině)
  - udává **jak** závisí střední hodnota **spojité** veličiny Y na nezávisle proměnné (proměnných) x
  - **nesymetrická** vlastnost (závislost Y na $x \neq$ závislost X na y)
  - lze použít k **prokazování** existence závislosti **závisle** proměnné Y na **nezávisle** proměnné x
  - umožňuje **předpovídat** stř. hodnotu Y pro zvolenou hodnotu x

korelace a regrese
[correlation, regression]

- **korelace** (dvojice náhodných veličin)
  - měří **sílu** (těsnost) **vzájemné** závislosti **spojitých** veličin
  - lze použít k **prokazování** existence **vzájemné** závislosti X, Y
  - k **porovnávání síly** (těsnosti) závislosti v několika populacích
  - **symetrická** vlastnost veličin X a Y
- **regrese** (náhodná veličina na nenáhodné veličině)
  - udává **jak** závisí střední hodnota **spojité** veličiny Y na nezávisle proměnné (proměnných) x
  - **nesymetrická** vlastnost (závislost Y na $x \neq$ závislost X na y)
  - lze použít k **prokazování** existence závislosti **závisle** proměnné Y na **nezávisle** proměnné x
  - umožňuje **předpovídat** stř. hodnotu Y pro zvolenou hodnotu x

korelace a regrese

[correlation, regression]

- **korelace** (dvojice náhodných veličin)
  - měří **sílu** (těsnost) **vzájemné** závislosti **spojitých** veličin
  - lze použít k **prokazování** existence **vzájemné** závislosti X, Y
  - k **porovnávání síly** (těsnosti) závislosti v několika populacích
  - **symetrická** vlastnost veličin X a Y
- **regrese** (náhodná veličina na nenáhodné veličině)
  - udává **jak** závisí střední hodnota **spojité** veličiny Y na nezávisle proměnné (proměnných) x
  - **nesymetrická** vlastnost (závislost Y na $x \neq$ závislost X na y)
  - lze použít k **prokazování** existence závislosti **závisle** proměnné Y na **nezávisle** proměnné x
  - umožňuje **předpovídat** stř. hodnotu Y pro zvolenou hodnotu x

korelace a regrese

[correlation, regression]

- **korelace** (dvojice náhodných veličin)
  - měří **sílu** (těsnost) **vzájemné** závislosti **spojitých** veličin
  - lze použít k **prokazování** existence **vzájemné** závislosti X, Y
  - k **porovnávání síly** (těsnosti) závislosti v několika populacích
  - **symetrická** vlastnost veličin X a Y
- **regrese** (náhodná veličina na nenáhodné veličině)
  - udává **jak** závisí střední hodnota **spojité** veličiny Y na nezávisle proměnné (proměnných) x
  - **nesymetrická** vlastnost (závislost Y na $x \neq$ závislost X na y)
  - lze použít k **prokazování** existence závislosti **závisle** proměnné Y na **nezávisle** proměnné x
  - umožňuje **předpovídat** stř. hodnotu Y pro zvolenou hodnotu x

korelace a regrese

[correlation, regression]

- **korelace** (dvojice náhodných veličin)
  - měří **sílu** (těsnost) **vzájemné** závislosti **spojitých** veličin
  - lze použít k **prokazování** existence **vzájemné** závislosti X, Y
  - k **porovnávání síly** (těsnosti) závislosti v několika populacích
  - **symetrická** vlastnost veličin X a Y
- **regrese** (náhodná veličina na nenáhodné veličině)
  - udává **jak** závisí střední hodnota **spojité** veličiny Y na nezávisle proměnné (proměnných) x
  - **nesymetrická** vlastnost (závislost Y na $x \neq$ závislost X na y)
  - lze použít k **prokazování** existence závislosti **závisle** proměnné Y na **nezávisle** proměnné x
  - umožňuje **předpovídat** stř. hodnotu Y pro zvolenou hodnotu x

korelace a regrese

[correlation, regression]

- **korelace** (dvojice náhodných veličin)
  - měří **sílu** (těsnost) **vzájemné** závislosti **spojitých** veličin
  - lze použít k **prokazování** existence **vzájemné** závislosti X, Y
  - k **porovnávání síly** (těsnosti) závislosti v několika populacích
  - **symetrická** vlastnost veličin X a Y
- **regrese** (náhodná veličina na nenáhodné veličině)
  - udává **jak** závisí střední hodnota **spojité** veličiny Y na nezávisle proměnné (proměnných) x
  - **nesymetrická** vlastnost (závislost Y na $x \neq$ závislost X na y)
  - lze použít k **prokazování** existence závislosti **závisle** proměnné Y na **nezávisle** proměnné x
  - umožňuje **předpovídat** stř. hodnotu Y pro zvolenou hodnotu x

korelace a regrese

[correlation, regression]

- **korelace** (dvojice náhodných veličin)
  - měří **sílu** (těsnost) **vzájemné** závislosti **spojitých** veličin
  - lze použít k **prokazování** existence **vzájemné** závislosti X, Y
  - k **porovnávání síly** (těsnosti) závislosti v několika populacích
  - **symetrická** vlastnost veličin X a Y
- **regrese** (náhodná veličina na nenáhodné veličině)
  - udává **jak** závisí střední hodnota **spojité** veličiny Y na nezávisle proměnné (proměnných) x
  - **nesymetrická** vlastnost (závislost Y na $x \neq$ závislost X na y)
  - lze použít k **prokazování** existence závislosti **závisle** proměnné Y na **nezávisle** proměnné x
  - umožňuje **předpovídat** stř. hodnotu Y pro zvolenou hodnotu x

korelační koeficient

(zavedení **výběrového** korelačního koeficientu)

- (populační) korelační koeficient $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ (zaveden na obr. 78)
  - $|\rho_{XY}| \leq 1$
  - pro nezávislé X, Y je $\rho_{XY} = 0$
  - měří sílu **lineární** závislosti
- (výběrový) korelační koeficient r_{xy} (zaveden na obr. 34)

$$r_{XY} = \frac{s_{xy}}{s_x s_y} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

  - odhaduje ρ_{XY}
  - přesnost odhadu závisí na n
- alternativní označení: **Pearsonův** korelační koeficient, momentový korelační koeficient, [correlation coefficient]

korelační koeficient

(zavedení **výběrového** korelačního koeficientu)

- (populační) korelační koeficient $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ (zaveden na obr. 78)
  - $|\rho_{XY}| \leq 1$
  - pro nezávislé X, Y je $\rho_{XY} = 0$
  - měří sílu **lineární** závislosti
- (výběrový) korelační koeficient r_{xy} (zaveden na obr. 34)

$$r_{XY} = \frac{s_{xy}}{s_x s_y} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

  - odhaduje ρ_{XY}
  - přesnost odhadu závisí na n
- alternativní označení: **Pearsonův** korelační koeficient, momentový korelační koeficient, [correlation coefficient]

korelační koeficient

(zavedení **výběrového** korelačního koeficientu)

- (populační) korelační koeficient $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ (zaveden na obr. 78)
  - $|\rho_{XY}| \leq 1$
  - pro nezávislé X, Y je $\rho_{XY} = 0$
  - měří sílu **lineární** závislosti
- (výběrový) korelační koeficient r_{xy} (zaveden na obr. 34)

$$r_{XY} = \frac{s_{xy}}{s_x s_y} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

  - odhaduje ρ_{XY}
  - přesnost odhadu závisí na n
- alternativní označení: **Pearsonův** korelační koeficient, momentový korelační koeficient, [correlation coefficient]

korelační koeficient

(zavedení **výběrového** korelačního koeficientu)

- (populační) korelační koeficient $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ (zaveden na obr. 78)
  - $|\rho_{XY}| \leq 1$
  - pro nezávislé X, Y je $\rho_{XY} = 0$
  - měří sílu **lineární** závislosti
- (výběrový) korelační koeficient r_{xy} (zaveden na obr. 34)

$$r_{XY} = \frac{s_{xy}}{s_x s_y} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

  - odhaduje ρ_{XY}
  - přesnost odhadu závisí na n
- alternativní označení: **Pearsonův** korelační koeficient, momentový korelační koeficient, [correlation coefficient]

korelační koeficient

(zavedení **výběrového** korelačního koeficientu)

- (populační) korelační koeficient $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ (zaveden na obr. 78)
  - $|\rho_{XY}| \leq 1$
  - pro nezávislé X, Y je $\rho_{XY} = 0$
  - měří sílu **lineární** závislosti
- (výběrový) korelační koeficient r_{xy} (zaveden na obr. 34)

$$r_{XY} = \frac{s_{xy}}{s_x s_y} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

  - odhaduje ρ_{XY}
  - přesnost odhadu závisí na n
- alternativní označení: **Pearsonův** korelační koeficient, momentový korelační koeficient, [correlation coefficient]

korelační koeficient

(zavedení **výběrového** korelačního koeficientu)

- (populační) korelační koeficient $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ (zaveden na obr. 78)
  - $|\rho_{XY}| \leq 1$
  - pro nezávislé X, Y je $\rho_{XY} = 0$
  - měří sílu **lineární** závislosti
- (výběrový) korelační koeficient r_{xy} (zaveden na obr. 34)

$$r_{XY} = \frac{s_{xy}}{s_x s_y} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

  - odhaduje ρ_{XY}
  - přesnost odhadu závisí na n
- alternativní označení: **Pearsonův** korelační koeficient, momentový korelační koeficient, [correlation coefficient]

korelační koeficient

(zavedení **výběrového** korelačního koeficientu)

- (populační) korelační koeficient $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ (zaveden na obr. 78)
  - $|\rho_{XY}| \leq 1$
  - pro nezávislé X, Y je $\rho_{XY} = 0$
  - měří sílu **lineární** závislosti
- (výběrový) korelační koeficient r_{xy} (zaveden na obr. 34)

$$r_{XY} = \frac{s_{xy}}{s_x s_y} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

  - odhaduje ρ_{XY}
  - přesnost odhadu závisí na n
- alternativní označení: **Pearsonův** korelační koeficient, momentový korelační koeficient, [correlation coefficient]

korelační koeficient

(zavedení **výběrového** korelačního koeficientu)

- (populační) korelační koeficient $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ (zaveden na obr. 78)
  - $|\rho_{XY}| \leq 1$
  - pro nezávislé X, Y je $\rho_{XY} = 0$
  - měří sílu **lineární** závislosti
- (výběrový) korelační koeficient r_{xy} (zaveden na obr. 34)

$$r_{XY} = \frac{s_{xy}}{s_x s_y} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

  - odhaduje ρ_{XY}
  - přesnost odhadu závisí na n
- alternativní označení: **Pearsonův** korelační koeficient, momentový korelační koeficient, [correlation coefficient]

dokazování závislosti X, Y

- k prokázání závislosti nutno **normální** rozdělení (X, Y)
- $H_0 : \rho_{XY} = 0$ se na hladině α zamítá:

$$T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}, \qquad |T| \geq t_{n-2}(\alpha)$$

  (r je dost daleko od nuly)
- **Spearmanův** korelační koeficient
  - měří sílu **monotonní** závislosti
  - založen na **pořadích** R_i, Q_i hodnot X_i, Y_i

$$r_{XY}^{(S)} = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^{n}(R_i - Q_i)^2$$

  - k testu nezávislosti nepotřebuje normální rozdělení
  - H_0: (nezávislost) se zamítá, je-li $|r_{XY}^{(S)}\sqrt{n-1}| \geq z(\alpha/2)$

dokazování závislosti X, Y

- k prokázání závislosti nutno **normální** rozdělení (X, Y)
- $H_0 : \rho_{XY} = 0$ se na hladině α zamítá:

$$T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}, \qquad |T| \geq t_{n-2}(\alpha)$$

(r je dost daleko od nuly)

- **Spearmanův** korelační koeficient
  - měří sílu **monotonní** závislosti
  - založen na **pořadích** R_i, Q_i hodnot X_i, Y_i

$$r_{XY}^{(S)} = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^{n}(R_i - Q_i)^2$$

  - k testu nezávislosti nepotřebuje normální rozdělení
  - H_0: (nezávislost) se zamítá, je-li $|r_{XY}^{(S)}\sqrt{n-1}| \geq z(\alpha/2)$

dokazování závislosti X, Y

- k prokázání závislosti nutno **normální** rozdělení (X, Y)
- $H_0 : \rho_{XY} = 0$ se na hladině α zamítá:

$$T = \frac{r}{\sqrt{1 - r^2}} \sqrt{n - 2}, \qquad |T| \geq t_{n-2}(\alpha)$$

(r je dost daleko od nuly)

- **Spearmanův** korelační koeficient
  - měří sílu **monotonní** závislosti
  - založen na **pořadích** R_i, Q_i hodnot X_i, Y_i

$$r_{XY}^{(S)} = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^{n} (R_i - Q_i)^2$$

  - k testu nezávislosti nepotřebuje normální rozdělení
  - H_0: (nezávislost) se zamítá, je-li $|r_{XY}^{(S)} \sqrt{n - 1}| \geq z(\alpha/2)$

dokazování závislosti X, Y

- k prokázání závislosti nutno **normální** rozdělení (X, Y)
- $H_0 : \rho_{XY} = 0$ se na hladině α zamítá:

$$T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}, \qquad |T| \geq t_{n-2}(\alpha)$$

(r je dost daleko od nuly)

- **Spearmanův** korelační koeficient
  - měří sílu **monotonní** závislosti
  - založen na **pořadích** R_i, Q_i hodnot X_i, Y_i

$$r_{XY}^{(S)} = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^{n}(R_i - Q_i)^2$$

  - k testu nezávislosti nepotřebuje normální rozdělení
  - H_0: (nezávislost) se zamítá, je-li $|r_{XY}^{(S)}\sqrt{n-1}| \geq z(\alpha/2)$

dokazování závislosti X, Y

- k prokázání závislosti nutno **normální** rozdělení (X, Y)
- $H_0 : \rho_{XY} = 0$ se na hladině α zamítá:

$$T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}, \qquad |T| \geq t_{n-2}(\alpha)$$

(r je dost daleko od nuly)

- **Spearmanův** korelační koeficient
  - měří sílu **monotonní** závislosti
  - založen na **pořadích** R_i, Q_i hodnot X_i, Y_i

$$r_{XY}^{(S)} = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} (R_i - Q_i)^2$$

  - k testu nezávislosti nepotřebuje normální rozdělení
  - H_0: (nezávislost) se zamítá, je-li $|r_{XY}^{(S)}\sqrt{n-1}| \geq z(\alpha/2)$

dokazování závislosti X, Y

- k prokázání závislosti nutno **normální** rozdělení (X, Y)
- $H_0 : \rho_{XY} = 0$ se na hladině α zamítá:

$$T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}, \qquad |T| \geq t_{n-2}(\alpha)$$

(r je dost daleko od nuly)

- **Spearmanův** korelační koeficient
  - měří sílu **monotonní** závislosti
  - založen na **pořadích** R_i, Q_i hodnot X_i, Y_i

$$r_{XY}^{(S)} = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^{n}(R_i - Q_i)^2$$

  - k testu nezávislosti nepotřebuje normální rozdělení
  - H_0: (nezávislost) se zamítá, je-li $|r_{XY}^{(S)}\sqrt{n-1}| \geq z(\alpha/2)$

dokazování závislosti X, Y

- k prokázání závislosti nutno **normální** rozdělení (X, Y)
- $\mathrm{H}_0 : \rho_{XY} = 0$ se na hladině α zamítá:

$$T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}, \qquad |T| \geq t_{n-2}(\alpha)$$

  (r je dost daleko od nuly)
- **Spearmanův** korelační koeficient
  - měří sílu **monotonní** závislosti
  - založen na **pořadích** R_i, Q_i hodnot X_i, Y_i

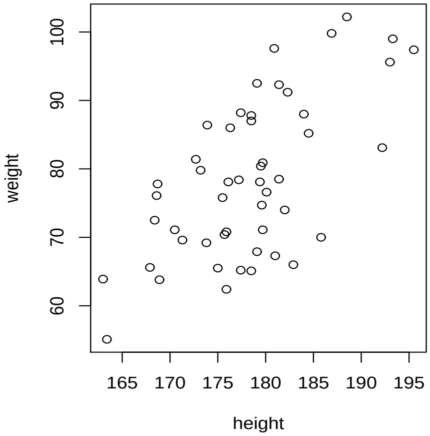
$$r_{XY}^{(S)} = 1 - \frac{6}{n(n^2-1)}\sum_{i=1}^{n}(R_i - Q_i)^2$$

  - k testu nezávislosti nepotřebuje normální rozdělení
  - H_0: (nezávislost) se zamítá, je-li $|r_{XY}^{(S)}\sqrt{n-1}| \geq z(\alpha/2)$

závislost váhy na výšce u mužů

data: Policie

[plot(weight~height)]



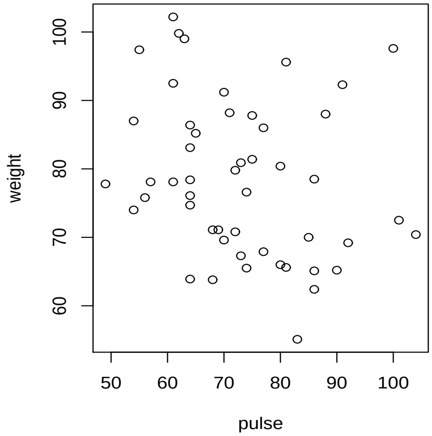
[cor.test(weight,height)]

$$r = 0{,}648$$
$$t = 5{,}814$$
$$p < 0{,}001$$

závislost váhy na pulsu u mužů

data: Policie

[plot(weight∼pulse)]



[cor.test(pulse,weight)]

$$r = -0{,}245$$
$$t = -1{,}752$$
$$p = 8{,}6\ \%$$