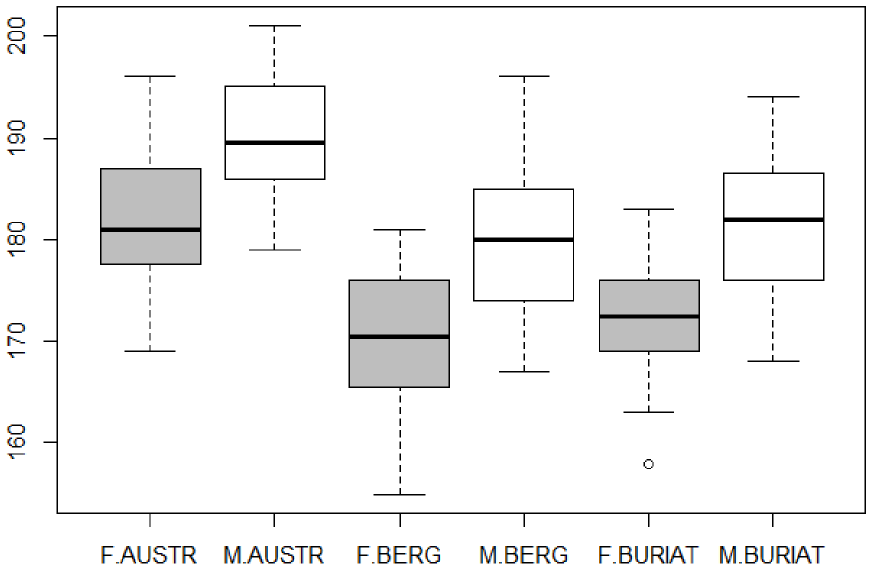
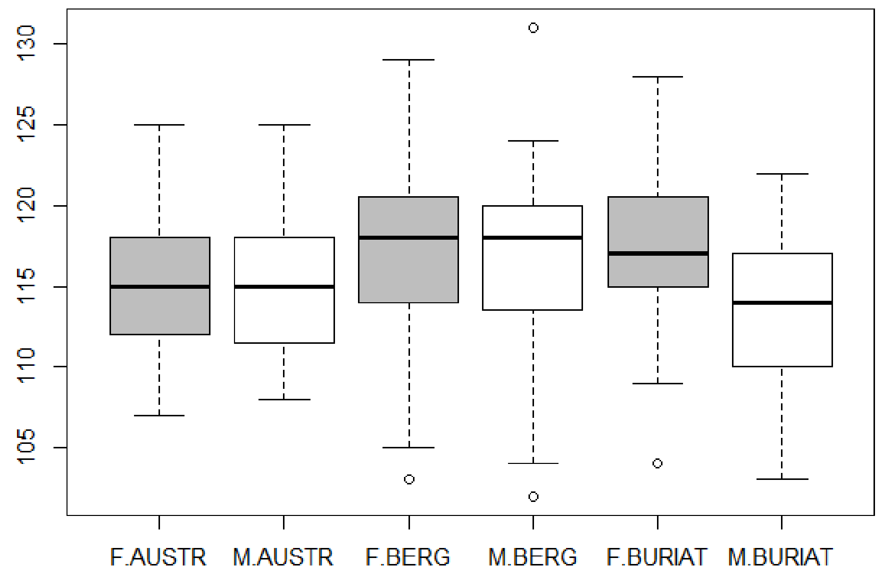
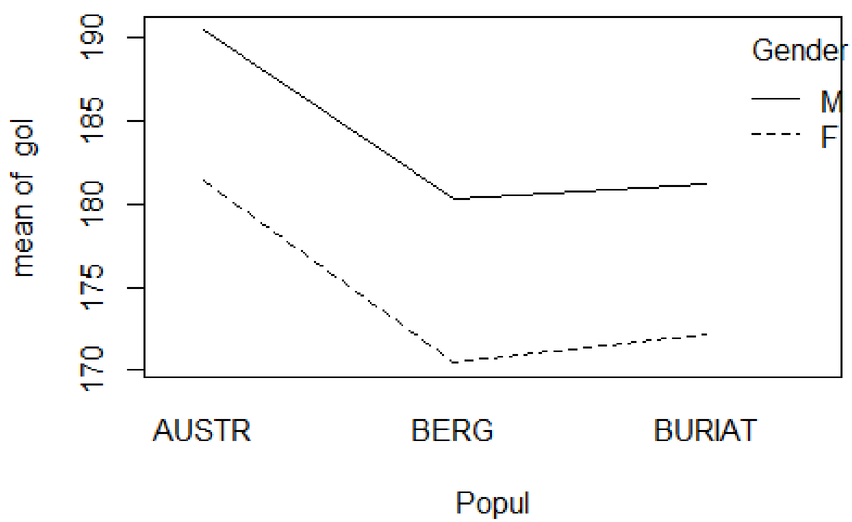
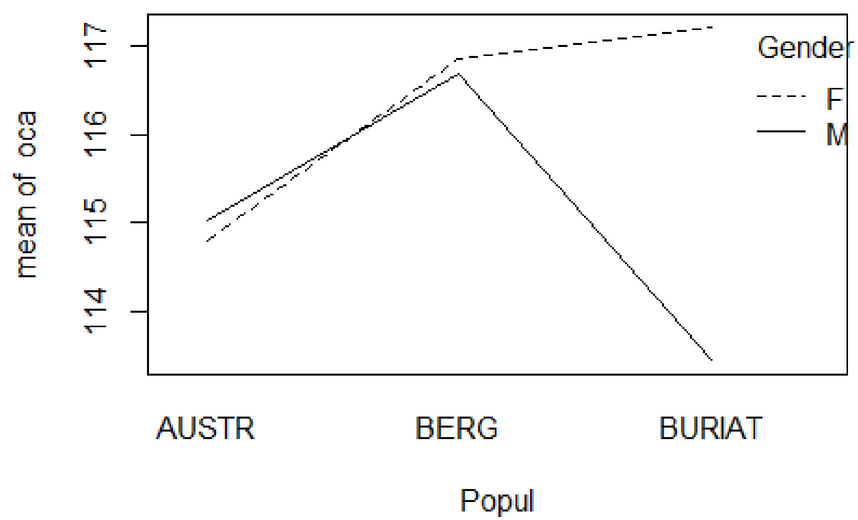
Příkladová data



Týlní úhel (oca)







Dvoucestná analýza rozptylu, analýza dvojného třídění

[two-way ANOVA, two-faktor ANOVA]

Model jednocestné analýzy:

Minule jsme pracovali s tímto modelem: (máme ***k*** výběrů)

$\boldsymbol{X_{it} = \mu_i + E_{it}}$ … hodnota $\boldsymbol{X_{it}}$ je dána střední hodnotou dané skupiny ***i*** a přirozenou variabilitou jedinců $\boldsymbol{E_{it}}$, kterou modelujeme jako náhodnou veličinu $\boldsymbol{E \sim N(0, \sigma^2)}$; ***i*** = 1, …, *k*; ***t*** = 1, …, n_i.

Jiný způsob zápisu, který budeme dále rozvíjet:

$\boldsymbol{X_{it} = \mu + \alpha_i + E_{it}}$ … hodnota $\boldsymbol{X_{it}}$ je dána společnou střední hodnotou $\boldsymbol{\mu}$ **+** odchylkou $\boldsymbol{\alpha_i}$ dané skupiny od společné stř. hodnoty **+** přirozenou variabilitou jedinců $\boldsymbol{E_{it} \sim N(0, \sigma^2)}$

Zde modelujeme ***k*** středních hodnot pomocí ***k+1*** parametrů ($\boldsymbol{\mu, \alpha_1, \ldots, \alpha_k}$). Tomu se říká *přeparametrizovaný model*. Aby bylo možné převést „staré“ parametry na „nové“ jednoznačně, přidává se tzv. **reparametrizační podmínka**, typicky

$$\sum_{i=1}^{k} \boldsymbol{\alpha_i = 0} \quad \text{nebo} \quad \boldsymbol{\alpha_1 = 0} \quad \text{(v Rku).}$$

Odhady těchto parametrů pak počítáme jako $\boldsymbol{\hat{\mu} = \bar{X}_{\bullet\bullet}}$, $\boldsymbol{\hat{\alpha}_i = \bar{X}_{\bullet\bullet} - \bar{X}_{i\bullet}}$.

Populační průměr $\boldsymbol{\bar{X}_i = \mu + \alpha_i}$.

Dvoucestná analýza rozptylu, analýza dvojného třídění

Model dvoucestné analýzy bez interakcí:

Rozšíříme zápis na dva faktory, **A** s úrovněmi 1, …, ***a*** a **B** s úrovněmi 1, …, ***b***:

$$\boldsymbol{X_{ijt} = \mu + \alpha_i + \beta_j + E_{ijt}}$$

… hodnota $\boldsymbol{X_{ijt}}$ je dána společnou střední hodnotou $\boldsymbol{\mu}$ **+** vlivem ***i***-té úrovně faktoru **A** ($\boldsymbol{\alpha_i}$) **+** vlivem ***j***-té úrovně faktoru **B** ($\boldsymbol{\beta_j}$) **+** přirozenou variabilitou jedinců $\boldsymbol{E \sim N(0, \sigma^2)}$

reparametrizační podmínky:

$\boldsymbol{\sum_{i=1}^{a} \alpha_i = 0}$ nebo (v Rku) $\boldsymbol{\alpha_1 = 0}$.

$\boldsymbol{\sum_{j=1}^{b} \beta_j = 0}$ nebo (v Rku) $\boldsymbol{\beta_1 = 0}$.

Předpoklady:

- Nezávislost pozorování mezi náhodnými bloky (vysvětlíme za chvíli)
- Shodnost rozptylů všech skupin
- Normalita reziduí $\left(X_{ijt} - \bar{X}_{ij\bullet}\right)$
- **Aditivita vlivu obou faktorů** (sčitatelnost) = účinek $\mathbf{\beta_j}$ úrovně ***j*** faktoru **B** se přičítá stejně ke všem skupinám úrovně ***i*** faktoru **A**. To také znamená, že vliv jednoho faktoru nezávisí na současné úrovni druhého faktoru.

Model dvoucestné analýzy bez interakcí [main effects ANOVA]

$$X_{ijt} = \mu + \alpha_i + \beta_j + E_{ijt}$$

Hypotézy:

1) Faktor **A**: není rozdíl mezi účinkem jednotlivých úrovní

$$H_{0A}\colon\ \alpha_1 = \alpha_2 = \cdots = \alpha_a$$

1) Faktor **B**: není rozdíl mezi účinkem jednotlivých úrovní

$$H_{0B}\colon\ \beta_1 = \beta_2 = \cdots = \beta_b$$

Kdy použijeme?

- Když máme důvod považovat „interakce“ za součást přirozené variability E_{ijt}. Potom bude „interakční“ variabilita součástí reziduální sumy čtverců SS_E.
- Když máme model bez opakování, tj. jen jedno pozorování pro každou kombinaci úrovní faktorů A a B. V takovém případě totiž nelze interakce testovat. Místo toho bychom mohli provést Tukeyův test aditivity.

Model dvoucestné analýzy bez interakcí

$$\boldsymbol{X_{ijt}} = \boldsymbol{\mu} + \boldsymbol{\alpha_i} + \boldsymbol{\beta_j} + \boldsymbol{E_{ijt}} \qquad i = 1, \dots, a\,; \quad j = 1, \dots, b.$$

Tabulka dvoucestné analýzy rozptylu bez interakcí:

Zdroj variability	Suma čtverců	DF	Průměrný čtverec	F	p-hodnota
Faktor A	$SS_A = \sum_{i=1}^{a} \boldsymbol{n_i}(\bar{X}_{i\bullet} - \bar{X}_{\bullet\bullet})^2$	$df_A = a - 1$	SS_A/df_A	F_A	
Faktor B	$SS_B = \sum_{j=1}^{b} \boldsymbol{n_j}\left(\bar{X}_{\bullet j} - \bar{X}_{\bullet\bullet}\right)^2$	$df_B = b - 1$	SS_B/df_B	F_B	
Reziduály	$SS_{Err} = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(X_{ij} - \bar{X}_{ij}\right)^2$	$df_{Err} = N - a - b + 1$	$SS_{Err}/df_{Err} = S^2$	--	--
Celková variabilita	$SS_{TOT} = SS_A + SS_B + SS_{Err}$	$df_{TOT} = N - 1$	--	--	--

Reziduální $\boldsymbol{MS_{Err}}$ je odhadem společné variance $\boldsymbol{\sigma^2}$.
Poznámka: uvedené výpočty součtu čtverců platí jen pro vyvážený design!

Model dvoucestné analýzy bez interakcí

$$X_{ijt} = \mu + \alpha_i + \beta_j + E_{ijt}$$

Další postup jako při jednocestném modelu, jen každý faktor testuji zvlášť.

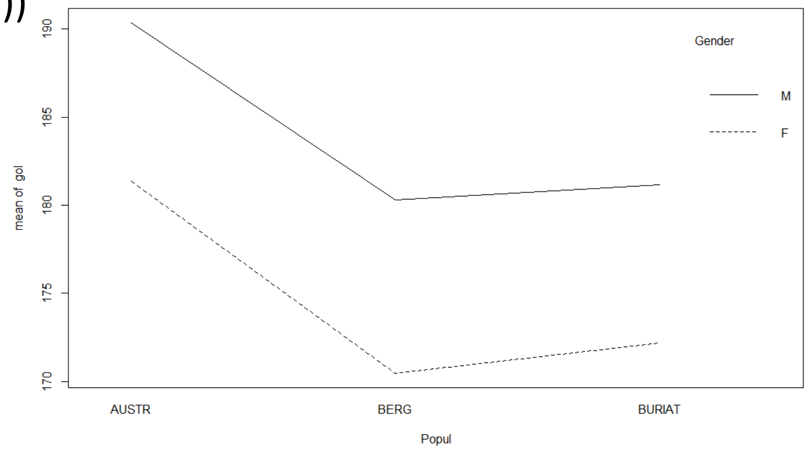
Za platnosti nulové hypotézy $\mathbf{H_{0A}}$ (resp. $\mathbf{H_{0B}}$) jsou $\mathbf{MS_A}$ (resp. $\mathbf{MS_B}$) malé a testová statistika je blízko nule:

$$F = \frac{MS_A}{MS_{Err}} = \frac{\frac{SS_A}{df_A}}{\frac{SS_{Err}}{df_{Err}}} \sim F_{df_A, df_{Err}}$$

Model dvoucestné analýzy bez interakcí

Příklad: nejdelší rozměr lebky ze tří populací: Austrálie, Rakousko a Burjatsko (Sibiř).

```
> summary(aGOL <- aov(gol~Gender*Popul,data=Howells))
             Df Sum Sq Mean Sq F value Pr(>F)
Gender        1   5171    5171  128.57 <2e-16 ***
Popul         2   5242    2621   65.17 <2e-16 ***
Gender:Popul  2     10       5    0.12  0.887
Residuals   234   9411      40
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```




```
> TukeyHSD(aGOL,which="Popul")
  Tukey multiple comparisons of means
    95% family-wise confidence level
diff       lwr       upr      p adj
BERG-AUSTR   -10.5 -12.865092 -8.134908 0.0000000 ***
BURIAT-AUSTR  -9.2 -11.565092 -6.834908 0.0000000 ***
BURIAT-BERG    1.3  -1.065092  3.665092 0.3985644    n.s.
```

Dvoucestná analýza rozptylu, analýza dvojného třídění

Model dvoucestné analýzy s interakcemi:

Dva faktory – **A** s úrovněmi 1, …, ***a*** a **B** s úrovněmi 1, …, ***b***:

$$\boldsymbol{X_{ijt} = \mu + \alpha_i + \beta_j + \gamma_{ij} + E_{ijt}}$$

… člen $\boldsymbol{\gamma_{ij}}$ popisuje variabilitu měření kolem očekávaného součtu vlivů úrovně $\boldsymbol{A_i}$ + úrovně $\boldsymbol{B_j}$. Dále testujeme, zda je tato variabilita významná (a je třeba ji zahrnout do modelu) nebo je zanedbatelná (neprůkazná).

reparametrizační podmínky:

$\sum_{i=1}^{a} \boldsymbol{\alpha_i = 0}$ nebo $\boldsymbol{\alpha_1 = 0}$ (v Rku) $\sum_{j=1}^{b} \boldsymbol{\beta_j = 0}$ nebo $\boldsymbol{\beta_1 = 0}$ (v Rku)

$\sum_{i=1}^{a} \boldsymbol{\gamma_{ij} = 0}$ $\sum_{j=1}^{b} \boldsymbol{\gamma_{ij} = 0}$

Předpoklady:

- Nezávislost pozorování mezi náhodnými bloky (vysvětlíme za chvíli)
- Shodnost rozptylů všech skupin
- Normalita reziduí $\left(X_{ijt} - \bar{X}_{ij\bullet}\right)$
- Aditivita v širším smyslu: model předpokládá, že vlivy (včetně interakčních) se sčítají. Jiný je model multiplikativní, kde se účinky násobí (vizte dále)

Model dvoucestné analýzy s interakcemi

$$\boldsymbol{X_{ijt} = \mu + \alpha_i + \beta_j + \gamma_{ij} + E_{ijt}}$$

Hypotézy:

1) Interakční členy jsou nulové: $\boldsymbol{\gamma_{ij} = 0}\ pro\ všechna\ \boldsymbol{i}\ a\ \boldsymbol{j}$. Pokud tuto hypotézu nezamítám, předpokládám čistě aditivní účinek faktorů **A** a **B** a testuji dále:
2) Faktor **A**: není rozdíl mezi účinkem jednotlivých úrovní $\boldsymbol{\alpha_1 = \alpha_2 = \cdots = \alpha_a}$
3) Faktor **B**: není rozdíl mezi účinkem jednotlivých úrovní $\boldsymbol{\beta_1 = \beta_2 = \cdots = \beta_b}$

Tabulka dvoucestné analýzy rozptylu s interakcemi:

Variabilita	**Suma čtverců**	**DF**	**Průměrný čtverec**	**F**	**p-hodnota**
Faktor A	SS_A	$df_A = a - 1$	SS_A/df_A	F_A	
Faktor B	SS_B	$df_B = b - 1$	SS_B/df_B	F_B	
Interakce A:B	SS_{AB}	$df_{AB} = (a-1)(b-1)$	SS_{AB}/df_{AB}	F_{AB}	
Reziduály	SS_{Err}	$df_{Err} = N - ab$	$SS_{Err}/df_{Err} = S^2$	--	--
celková	SS_{TOT}	$df_{TOT} = N - 1$	--	--	--

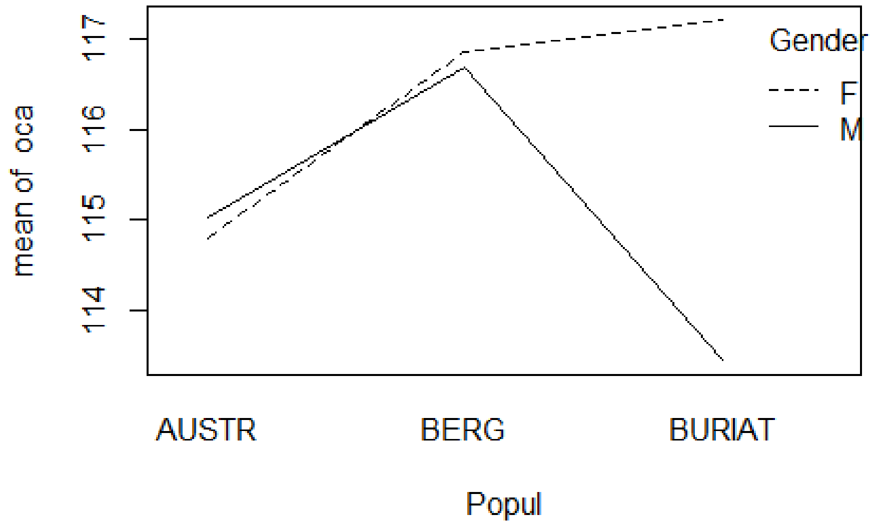
$$SS_{AB} = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{n_{ij}}{N}\left(\bar{X}_{ij} - \left(\bar{X}_{i\bullet} + \bar{X}_{\bullet j} - \bar{X}_{\bullet\bullet}\right)\right)^2$$

Model dvoucestné analýzy s interakcemi

Příklad: lebky – týlní úhel:

```
> summary(aOCA <- aov(oca~Gender*Popul,
data=Howells))

              Df Sum Sq Mean Sq F value Pr(>F)
Gender         1     91   91.27   3.689 0.0560 .
Popul          2    151   75.45   3.050 0.0493 *
Gender:Popul   2    192   95.80   3.872 0.0222 *
Residuals    234   5790   24.74
```



Když jsou interakce významné, studujeme dále všechny možné dvojice:
Jsme opatrní s interpretací vlivu samotných faktorů, protože při významných interakcích neplatí „čistá“ sčitatelnost účinků A + B.
(Přejít k modelu s jedním faktorem, který má a*b úrovní:)

```
> summary(aOCA_AB <- aov(oca~Gender:Popul,data=Howells))
              Df Sum Sq Mean Sq F value  Pr(>F)
Gender:Popul   5    434   86.76   3.507 0.00447 **
Residuals    234   5790   24.74
```

Model dvoucestné analýzy s interakcemi

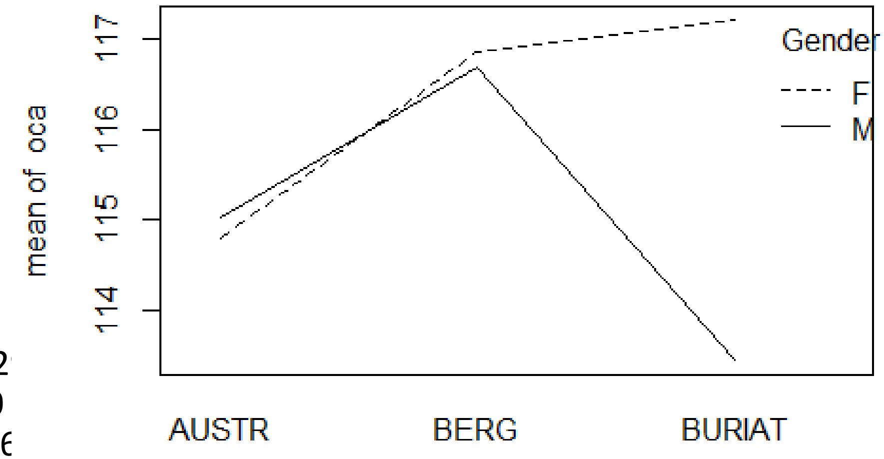
Příklad: lebky – týlní úhel:

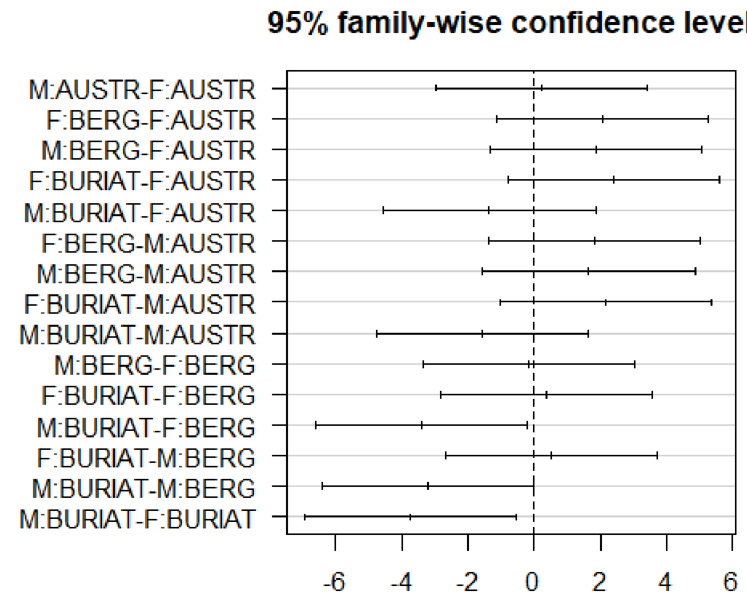
\> TukeyHSD(aOCA_AB)
Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = oca ~ Gender:Popul, data = Howells)

$`Gender:Popul`

	diff	lwr	upr	p adj
M:AUSTR-F:AUSTR	0.225	-2.9710152	3.42101522	0.999952
F:BERG-F:AUSTR	2.050	-1.1460152	5.24601522	0.4401940
M:BERG-F:AUSTR	1.875	-1.3210152	5.07101522	0.542568
F:BURIAT-F:AUSTR	2.400	-0.7960152	5.59601522	0.2617598
M:BURIAT-F:AUSTR	-1.350	-4.5460152	1.84601522	0.829852
F:BERG-M:AUSTR	1.825	-1.3710152	5.02101522	0.5723613
M:BERG-M:AUSTR	1.650	-1.5460152	4.84601522	0.6751737
F:BURIAT-M:AUSTR	2.175	-1.0210152	5.37101522	0.3712912
M:BURIAT-M:AUSTR	-1.575	-4.7710152	1.62101522	0.7172800
M:BERG-F:BERG	-0.175	-3.3710152	3.02101522	0.9999865
F:BURIAT-F:BERG	0.350	-2.8460152	3.54601522	0.9995849
M:BURIAT-F:BERG	-3.400	-6.5960152	-0.20398478	**0.0296922**
F:BURIAT-M:BERG	0.525	-2.6710152	3.72101522	0.9970591
M:BURIAT-M:BERG	-3.225	-6.4210152	-0.02898478	**0.0465369**
M:BURIAT-F:BURIAT	-3.750	-6.9460152	-0.55398478	**0.0111671**





Model dvoucestné analýzy bez opakování (a bez interakcí)

Pro každou kombinaci úrovní faktorů A a B (pro každou buňku) mám jen jedno pozorování.

Proto nelze spočítat průměr pro buňku $\bar{X}_{ij}$ ani modelovat interakce.

Rezidua počítám jako $\boldsymbol{X_{ij} - \bar{X}_{i\bullet} - \bar{X}_{\bullet j} + \bar{X}_{\bullet\bullet}}$

Reziduální součet čtverců **SS_E** bývá někdy označován jako „zbytkový“ součet čtverců [remainder SS], protože vzhledem k modelu s interakcemi se do něj „schová“ **SS_E + SS_{AB}**.

Dvoucestná analýza rozptylu

Friedmanův test pro model bez opakování (neparametrický)

- Pro faktor s pevným efektem v modelu bez opakování
- Pořadový test: mám-li testovaný faktor uspořádaný ve sloupcích, určuji pořadí zvlášť pro každý řádek.
- Myšlenka: Když $\mathbf{H_{0B}}$ platí (není rozdíl mezi úrovněmi faktoru B), budou pořadí v řádku náhodnou permutací čísel 1 až ***b***. Potom by součty pořadí ve sloupcích měly být srovnatelné.
- Testová statistika:

$$Q = \frac{12}{a \cdot b \cdot (b+1)} \cdot \sum_{j=1}^{b} \left(\sum_{i=1}^{a} R_{ij} \right)^2 - 3a(b+1) \sim_{asymptoticky} \chi^2_{b-1}$$

Pro malá ***a***, ***b*** tabulované přesné hodnoty.

Analýza rozptylu

Náhodný vs. pevný efekt faktoru [random vs. fixed effect]

Pevný efekt

- zajímají nás konkrétně zvolené úrovně faktoru;
- úrovně mohu zopakovat i v příštím experimentu;
- pro faktor s pevným efektem potřebuji reparametrizační podmínku;
- Př.: přídavek hnojiva, typ diety, rozdíl mezi druhy, mezi pohlavími

Náhodný efekt

- úroveň faktoru tvoří např. mláďata z jednoho vrhu, snímky z jedné lokality, paraziti z jedné ryby, apod.;
- celý faktor tedy rozlišuje různé vrhy mláďat, různé lokality, různé ryby, které ovšem chápeme jako náhodný výběr ze všech vrhů/lokalit/ryb daného typu.
- Primárně nás nezajímají rozdíly mezi měřenými vrhy/lokalitami/rybami, ale zajímá nás, zda se obecně vrhy/lokality/ryby mezi sebou liší.
- Úrovním faktoru s náhodným efektem pak říkáme **náhodné bloky**.

(Připouštíme ale i situaci, kdy nás zajímají právě ty zvolené vrhy/lokality/ryby. Potom je to faktor s pevnými efekty a řídí se pravidly pro pevné efekty.)

Model dvoucestné analýzy s náhodnými bloky

Symbolické znázornění náhodného bloku: A_i

$$X_{ijt} = \mu + A_i + \beta_j + \gamma_{ij} + E_{ijt}$$

- Říkáme tomu **smíšený model** [mixed effects model]
- Předpokládáme, že všechna A_i pochází z $N(0, \sigma_A^2)$
- Pro A_i nepotřebuji reparametrizační podmínku
- Nezávislost: v rámci jednoho náhodného bloku jsou pozorování závislá; nezávislost však musíme dodržet mezi náhodnými bloky (zajistím organizací experimentu).
- Hypotéza: mezi úrovněmi není žádná variabilita, tj. $\sigma_A^2 = 0$.
  (Netestujeme rovnost $A_1 = \cdots = A_a$)

Poznámka: Liší se přístupy k testování efektů ve smíšených modelech. Někteří autoři doporučují testovat pevný efekt (např. hnojení) proti interakčnímu MS (např. proti variabilitě odpovědi na hnojení na jednotlivých lokalitách): $F = \frac{MS_B}{MS_{AB}}$.
Více v učebnici Lepše a Šmilauera (str. 187, Tabulka 9-2).

Model dvoucestné analýzy
Ověření předpokladů

- Nezávislost
- Normalita reziduí
- Shodnost rozptylů

- **Aditivita faktorových účinků**
  a) Zkoumám graf závislosti hodnoty směrodatných odchylek (pro každou buňku) na příslušných průměrech. Pokud závislost existuje (a je pozitivní) , potom většinou pomůže logaritmická transformace.
  b) Je účinek interakcí průkazný? Potom nemohu model interpretovat jako „čistě“ aditivní, už nefunguje, že $\bar{X} = A + B$. Zaměříme se na interpretaci samotných interakcí.
  c) V modelu bez opakování nemohu testovat účinek interakcí, ale existuje **Tukeyho test aditivity**. Pro Rko máme skript: `TukeyTest(y, A, B)`
  Tukey navrhl modelování interakcí takto: $\gamma_{ij} = \gamma\alpha_i\beta_j$ a H_{0AB}: $\gamma = 0$.
  Z tabulky analýzy rozptylu upraveného modelu vyčtu rozhodnutí.

$$SS_{AB} = \frac{ab}{SS_A SS_B}\left[\sum_{i=1}^{a}\sum_{j=1}^{b} X_{ij}(\bar{X}_{i\bullet} - \bar{X}_{\bullet\bullet})(\bar{X}_{\bullet j} - \bar{X}_{\bullet\bullet})\right]^2, \qquad df_{AB} = 1.$$

Analýza rozptylu

Transformace dat – oprava násobného účinku faktorů

Předpokládáme sčitatelnost účinků faktorů:

$$\boldsymbol{X_{ijt} = \mu + \alpha_i + \beta_j\ (+\gamma_{ij}) + E_{ijt}}$$

Někdy je však **účinek multiplikativní**:

$$\boldsymbol{X_{ijt} = \mu * \alpha_i * \beta_j(*\,\gamma_{ij}) * E_{ijt}}$$

Projeví se mimo jiné tím, že s velikostí průměrných hodnot roste i směrodatná odchylka výběru.

Můžeme to vypozorovat také z poměrů průměrů ve skupinách.

Vymyšlený příklad:

	Faktor B		
Faktor A	*Úroveň B1*	*Úroveň B2*	*Úroveň B3*
Úroveň A1	10	30	60
Úroveň A2	20	60	120

→ Zavedeme **logaritmickou transformaci** (dále)

Dvoucestná analýza rozptylu

Transformace dat – přiblížení dat normálnímu rozdělení

Logaritmická transformace

$$\boldsymbol{Y_{ijt} = \log(X_{ijt})}$$

Kdy použít? Když mají faktory spíše násobný účinek (viz předchozí strana)

Když mají data výrazně pozitivně šikmé rozdělení.

Typické pro koncentrace látek, pro hmotnosti i rozměry, také počty jedinců, zvláště při shlukovitém rozmístění.

Omezení: data nesmí obsahovat *záporná čísla*.

Také *nuly jsou problém*, ale zde pomůže přidat posunutí: $Y_{ijt} = \log(X_{ijt} + \boldsymbol{c})$, kde ***c*** odpovídá nejmenším nenulovým hodnotám v datech, případně ***c*** = 1.

Převod na aditivitu:

$$\log\left(\boldsymbol{\mu} * \boldsymbol{\alpha_i} * \boldsymbol{\beta_j} * \boldsymbol{E_{ijt}}\right) = \log(\boldsymbol{\mu}) + \log(\boldsymbol{\alpha_i}) + \log(\boldsymbol{\beta_j}) + \log\left(\boldsymbol{E_{ijt}}\right)$$

Při použití posunutí ***+c*** platí aditivita jen přibližně.

Pro výsledný tvar rozdělení hodnot je jedno, zda použijeme přirozený *LN* nebo dekadický *LOG10* či jiný. Z praktického hlediska je výhodný dekadický logaritmus, protože *log*(10) = 1, *log*(100) = 2, atd.

Dvoucestná analýza rozptylu

Transformace dat

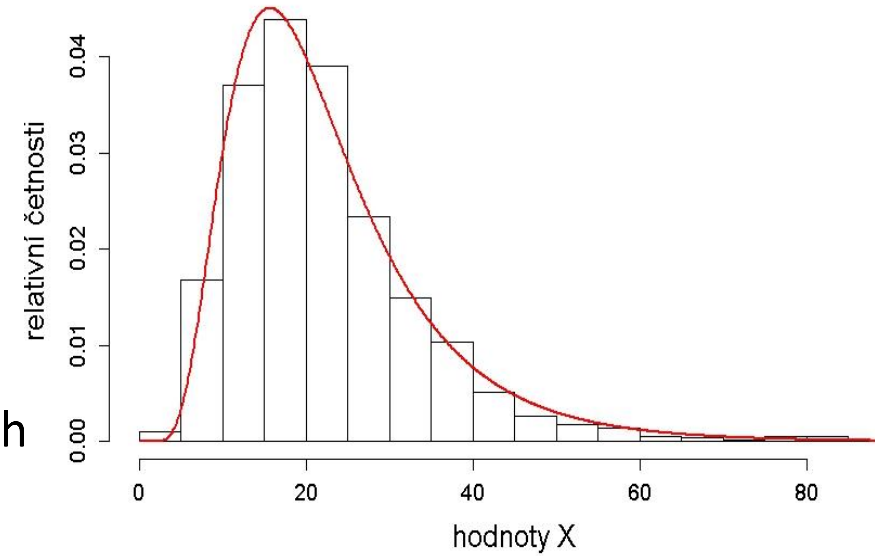
Logaritmická transformace

$$Y_{ijt} = \log(X_{ijt})$$

Připomenutí: když $\log(X) \sim N(\mu, \sigma^2)$, pak $X \sim LN(\mu, \sigma^2)$, logaritmicko-normální r.

Tři v jednom:

- odstraní násobný vztah faktorů
- odstraní závislost μ a σ
- přiblíží rozdělení dat normálnímu rozd.



Průměr a konfidenční interval:

Mohu použít hodnoty ze zlogaritmovaných výpočtů, které zpátky „odlogaritmuji“

$e^Y \rightarrow$ původní škála X hodnot

Průměr ale bude trochu vychýlený a konfidenční interval asymetrický (což je v pořádku, původní rozdělení dat je také asymetrické).

Dvoucestná analýza rozptylu

Transformace dat

Arcsinová transformace, také **angulární**

$$\boldsymbol{Y_{ijt} = arcsin\sqrt{p_{ijt}}}$$

R: **`asin(X)`**

Pro data o podílech či procentech.

Doporučeno pro $p < 0.3$ a $p > 0.7$.
Jsou-li hodnoty $p \in \langle 0.3,\ 0.7 \rangle$, není třeba transformovat, rozdělení bude dostatečné blízké normálnímu rozdělení.

Dvoucestná analýza rozptylu

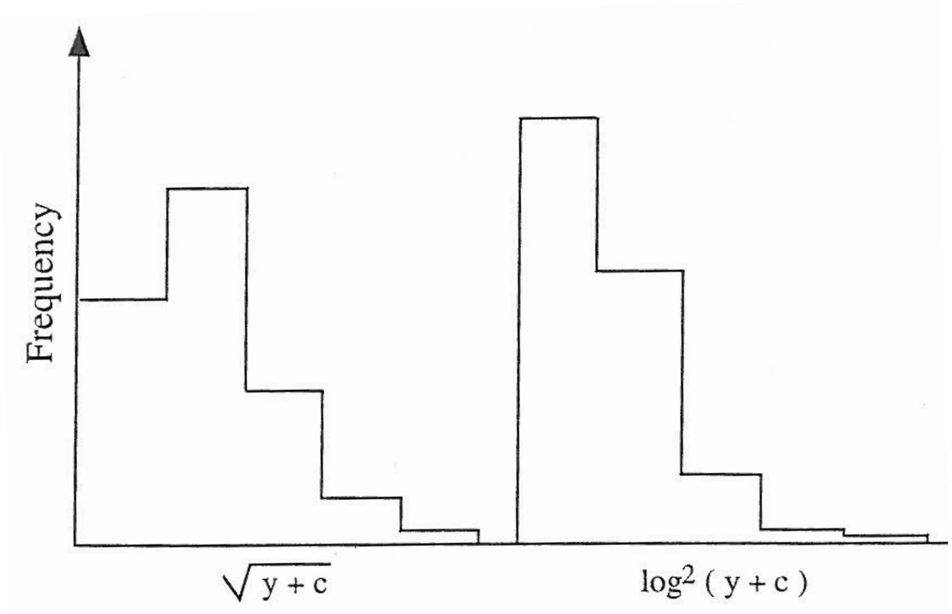
Transformace dat

Odmocninová transformace

$Y = \sqrt{X}$, také $Y = \sqrt{X + 0.5}$ pokud data obsahují nuly

Typicky pro data odpovídající Poissonovu rozdělení (funguje i log transform.), tedy počty jedinců v objemové či časové jednotce, pokud jsou rozmístěni náhodně.

Boxova-Coxova transformace (STAT)
Vylepšená forma mocninné transformace.



Analýza rozptylu – opakování z předchozí přednášky

Uspořádání experimentu [design of the experiment]

Faktoriálové uspořádání = design zahrnuje všechny kombinace úrovní faktorů (například data „npk“ přímo v Rku.)

Vyvážené uspořádání [balanced design]

- Ideálně stejný počet pozorování pro všechny kombinace úrovní faktorů (pro všechny buňky)
- S nevyvážeností počtu pozorování klesá síla testů.
- Proporční uspořádání = varianta, když není možné dodržet vyvážené uspořádání

$$n_{ij} = \frac{(\text{počet pozorov. s úrovní } \boldsymbol{i} \text{ faktoru } \boldsymbol{A})(\text{počet pozorov. s úrovní } \boldsymbol{j} \text{ faktoru } \boldsymbol{B})}{N} =$$

$$= \frac{\sum_{j=1}^{b} n_{ij} \cdot \sum_{i=1}^{a} n_{ij}}{N}$$

	B1	B2	…	Bj	…	Bb
A1				n_{1j}		
Ai	n_{i1}	n_{i2}	$\cdots$	$\boldsymbol{n_{ij}}$	$\cdots$	$n_{i\boldsymbol{b}}$
…				$\vdots$		
Aa				$n_{\boldsymbol{a}j}$		

Analýza rozptylu

Uspořádání experimentu

Nevyvážený design:

- přestává platit vztah $SS_{TOT} = SS_A + SS_B + SS_{AB} + SS_{Err}$
- klesají síly testů
- existují různé výpočtové metody rozkladu vysvětlené variability mezi jednotlivé efekty (faktory). Výsledky analýz z různých softwarů se proto mohou hodně lišit.

Analýza rozptylu

Uspořádání experimentu

Kde studovat více o typech rozkladu variability na součty čtverců:

- začněte v Lepš & Šmilauer, str. 219. Dále STATISTICA help a další anglické učebnice.
- Základní R-ková funkce **`aov()`** používá první typ rozkladu, tzv. sekvenciální rozklad. V knihovně **`car`** je dostupná funkce **`Anova()`**, která nabízí druhý a třetí typ rozkladu (parciální a ortogonální rozklad). Například pro model se 2 pevnými efekty, který má nesignifikantní interakce doporučuje Lepš a Šmilauer použít právě typ rozkladu II.

Analýza rozptylu - **Uspořádání experimentu**

Rozmístění pokusných ploch v terénu

Biotické i abiotické charakteristiky se v ploše kontinuálně mění => blízké plochy jsou si podobnější než plochy vzdálené. Proto byly vytvořeny standardní postupy pro uspořádání pokusů. Cílem je co nejvíce omezit vliv heterogenity pokusné plochy při zachování nezávislosti opakování.

Latinský čtverec
Každý řádek a každý sloupec obsahuje právě jednu plochu od každého typu zásahu.
Analyzujeme 3cestným modelem ANOVA
bez opakování.
Pro každou plochu zaznamenám řádek (faktor A), sloupec (B) a typ zásahu (C).

columns

rows

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

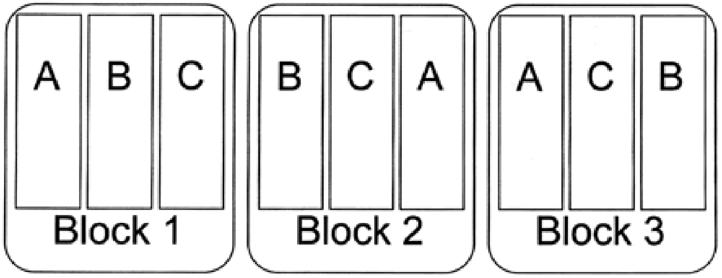
Variantou je také <u>úplně znáhodněné uspořádání</u>: typ zásahu na plochu určujeme náhodně (generátorem náhodných čísel) tak, aby každá plocha měla stejnou pravděpodobnost dostat kterýkoli druh zásahu.

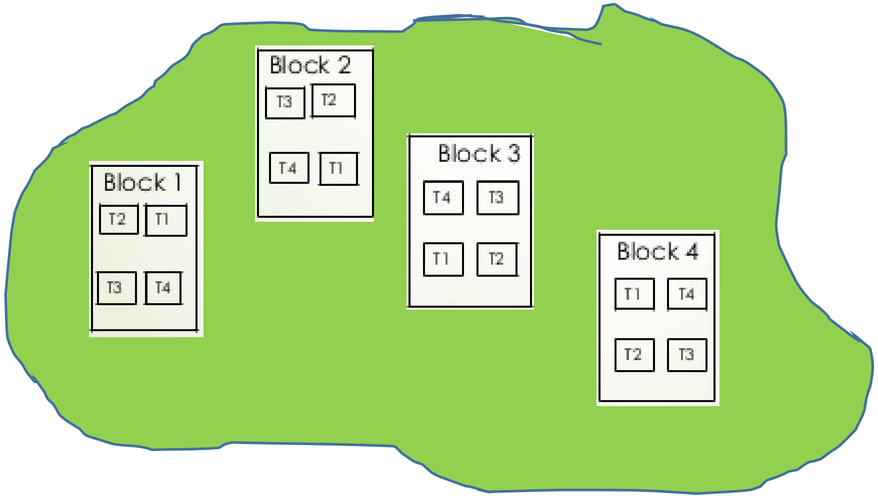
Analýza rozptylu - **Uspořádání experimentu**

Rozmístění pokusných ploch v terénu

Úplné znáhodněné bloky [randomized completely blocks]

- Výhodné pro velkou plochu. Potřebný počet opakování určuje počet bloků. Umístění bloků je vybíráno tak, aby byly vnitřně maximálně homogenní a zároveň aby odlišnosti mezi bloky pokrývaly variabilitu podmínek, přes které chceme naše závěry zobecňovat.
- Toto uspořádání umožňuje dobře odlišit prostorovou variabilitu od vlivu experimentálního zásahu.
- Pro homogenní plochu (rozdíly mezi bloky minimální) se naopak snižuje síla testu, zvláště, když je počet opakování malý.





Analýza rozptylu - **Uspořádání experimentu**

Další typ uspořádání

Opakovaná měření [repeated measures]

- Situace, kdy na jednom subjektu opakovaně měřím/hodnotím stejnou charakteristiku.
- např. přírůstky na rostlině či zvířeti; reakce jedince na různé podněty: typ stravy, chuť vína, alergeny;
- Subjekt je náhodný blok.
- Pozor na situaci, kdy jeden zásah může ovlivnit výsledek dalšího zásahu (např. při posuzování chuti vína není jedno, jestli jde suché víno po sladkém nebo naopak). Toto je třeba ošetřit např. dostatečným časovým odstupem. Nežádoucí vliv se dá snížit také větším počtem pokusných subjektů a náhodným pořadím provedených zásahů.

Vlastní uspořádání

Plánujete-li experiment s nestandardním uspořádáním, vždy nejdřív plán konzultujte se školitelem. Může se totiž stát, že data nepůjdou vyhodnotit žádnou statistickou metodou, nebo místo opakování naměříte tzv. *pseudoreplikace*.