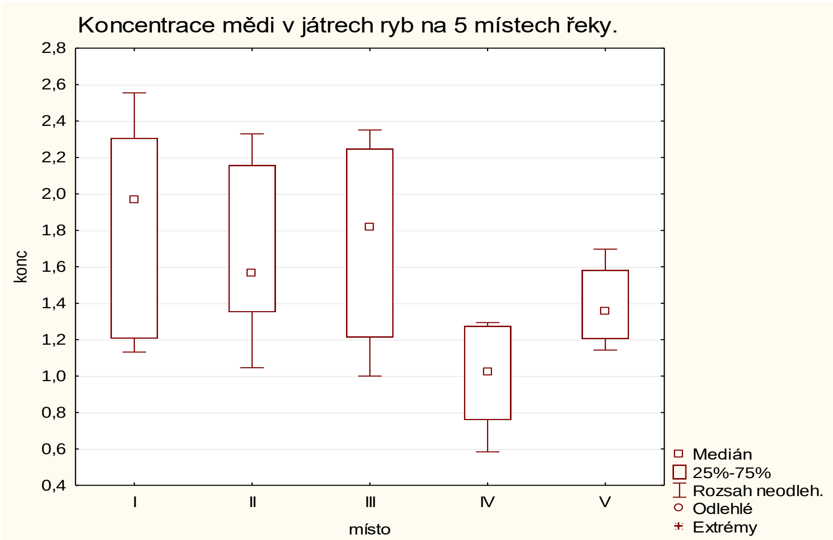
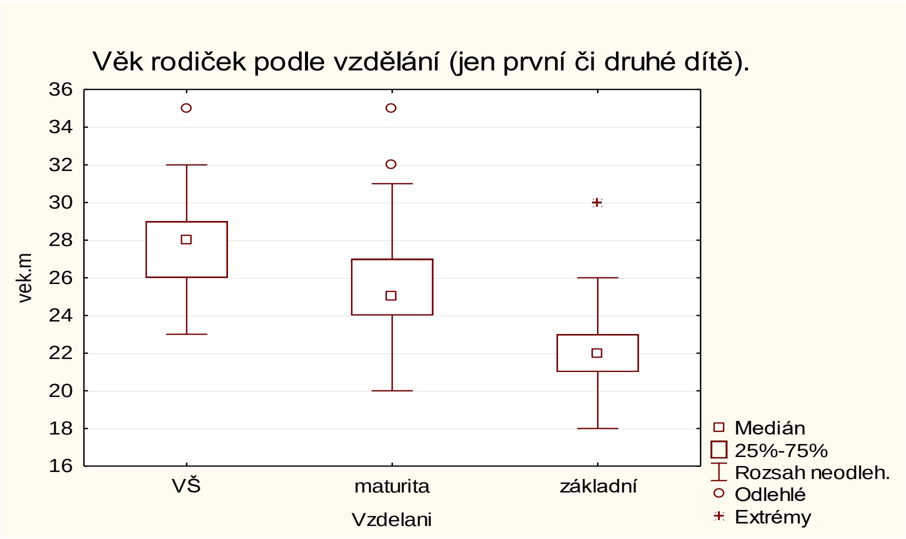
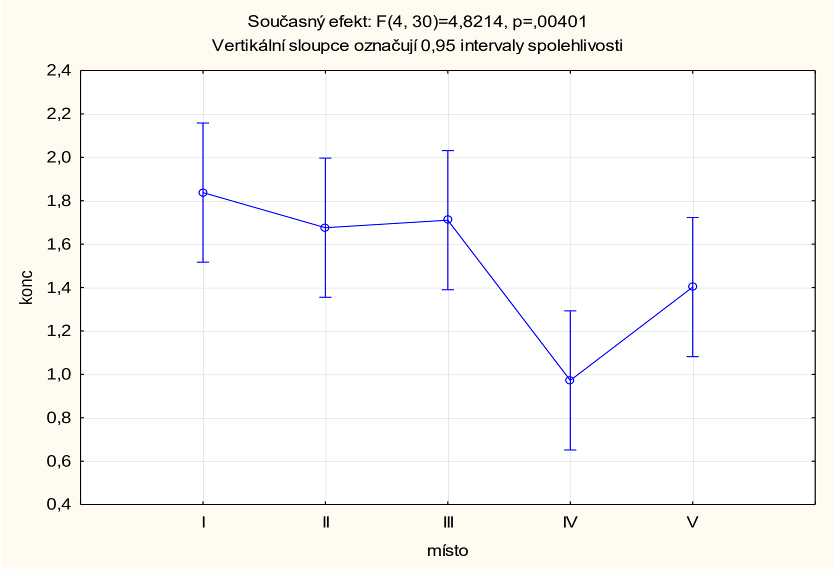
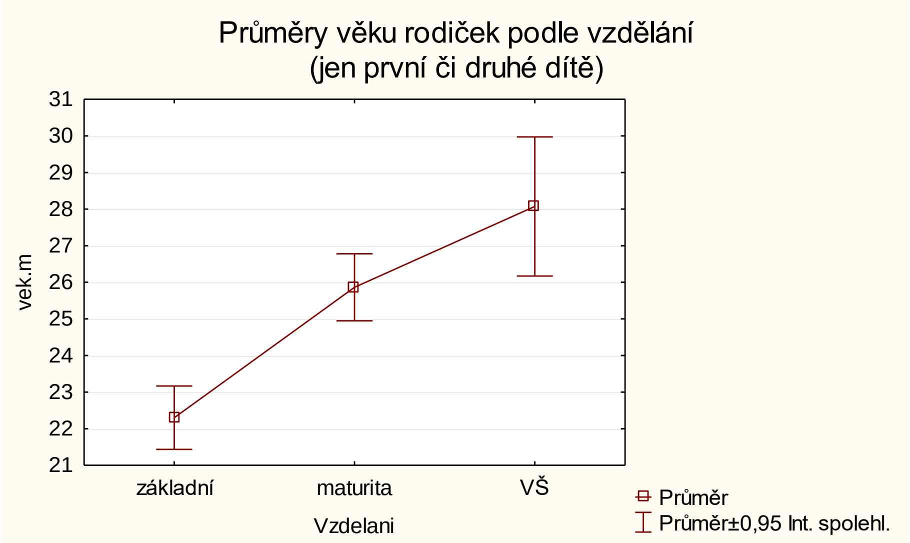
Několik výběrů tříděných podle jednoho faktoru









Analýza rozptylu, analýza variance [analysis of variance, ANOVA]

Mám několik skupin (výběrů) a ptám se, **zda jsou jejich průměry srovnatelné**?

Hypotéza: $\boldsymbol{H_0: \mu_1 = \mu_2 = \cdots = \mu_k}$ střední hodnoty všech $\boldsymbol{k}$ skupin jsou stejné

Alternativa $\boldsymbol{H_1}$: alespoň jedna střední hodnota se významně liší od ostatních

Myšlenka: jsou-li průměry skupin srovnatelné, potom bude *variabilita mezi skupinovými průměry* podstatně menší než *variabilita uvnitř skupin.*

Analýza tedy pracuje s odhadem rozptylů a testuje je pomocí F-testu.

Poznámka: Není správné testovat vztahy po dvojicích pomocí t-testů, protože velmi narůstá chyba prvního druhu:

Počet průměrů (k)	Hladina signifikance užívaná v *t*-testech			
	0.05	**0.02**	**0.01**	**0.001**
2	0.05	0.02	0.01	0.001
3	0.13	0.05	0.03	0.003
4	0.21	0.09	0.05	0.006
5	0.23	0.13	0.07	0.009
10	0.63	0.37	0.23	0.034
20	0.92	0.71	0.52	0.109
∞	1.00	1.00	1.00	1.00

Nejjednodušší situace:

Jednocestná analýza rozptylu, analýza jednoduchého třídění

[one-way ANOVA, single faktor ANOVA]

Skupiny jsou rozlišené **jedním faktorem**, např. příslušnost ke druhu. Jednotlivé druhy pak představují hladiny faktoru [levels of the factor].

Zároveň nás zajímají právě ty konkrétní druhy, jedná se o model s pevnými efekty [fixed effect model]. Odlišný model s *náhodnými efekty* popíšeme později.

Předpoklady testu:

- **Skupiny jsou náhodnými, vzájemně nezávislými výběry.** Tomu se říká zcela náhodné uspořádání [completely randomized experimental design]. Jiná uspořádání zahrnují například opakovaná měření na stejném jedinci, mláďata z jednoho vrhu, vegetační snímky z jedné lokality apod.
- **Každý výběr pochází z normálního rozdělení** $N\left(\mu_i, {\sigma^2}_i\right)$
- **Homoskedasticita, tj. variance výběrů jsou stejné:** ${\sigma^2}_1 = {\sigma^2}_2 = \cdots = {\sigma^2}_k$

Jednocestná analýza rozptylu - značení

Značení:

1. skupina: $(X_{11}, X_{12}, X_{13}, \dots, X_{1n_1})$ $\quad \bar{X}_{1\bullet} = \frac{1}{n_1}\sum_{t=1}^{n_1} X_{1t}$ $\quad \boldsymbol{S_1^2 = \frac{\sum_{t=1}^{n_1}(X_{1t} - \bar{X}_{1\bullet})^2}{n_1 - 1}}$

2. skupina: $(X_{21}, X_{22}, X_{23}, \dots, X_{2n_2})$ $\quad \bar{X}_{2\bullet} = \frac{1}{n_2}\sum_{t=1}^{n_2} X_{2t}$ $\quad \boldsymbol{S_2^2 = \frac{\sum_{t=1}^{n_2}(X_{2t} - \bar{X}_{2\bullet})^2}{n_2 - 1}}$

$\vdots$

$\boldsymbol{k}$. skupina: $(X_{k1}, X_{k2}, X_{k3}, \dots, X_{kn_k})$ $\quad \bar{X}_{k\bullet} = \frac{1}{n_k}\sum_{t=1}^{n_k} X_{kt}$ $\quad \boldsymbol{S_k^2 = \frac{\sum_{t=1}^{n_k}(X_{kt} - \bar{X}_{k\bullet})^2}{n_k - 1}}$

$\boldsymbol{N = n_1 + n_2 + \dots + n_k}$ … celkový počet pozorování;
počty pozorování ve skupinách se mohou lišit

$$\bar{X}_{\bullet\bullet} = \frac{1}{N}\sum_{i=1}^{k}\sum_{t=1}^{n_i} X_{it} = \frac{1}{N}\sum_{i=1}^{k} n_i \cdot \bar{X}_{i\bullet} = \boldsymbol{\sum_{i=1}^{k} \frac{n_i}{N} \cdot \bar{X}_{i\bullet}}$$

Poměrná část, součet vážených průměrů

Reziduály neboli odchylky od skupinových průměrů: $X_{it} - \bar{X}_{i\bullet}$

Jednocestná analýza rozptylu – sumy čtverců

Variabilita uvnitř skupin neboli reziduální rozptyl:

Suma čtverců reziduálů: [error sum of squares, residual SS]

$$SS_E = \sum_{i=1}^{k}\sum_{t=1}^{n_i}(X_{it} - \bar{X}_{i\bullet})^2$$

$$DF_E = N - k$$

Průměrný čtverec odchylky od skupinového průměru: [within group mean square, error mean square, residual MS]

$$MS_E = \frac{SS_E}{N-k} = \frac{\sum_{i=1}^{k}\sum_{t=1}^{n_i}(X_{it} - \bar{X}_{i\bullet})^2}{\sum_{i=1}^{k}(n_i - 1)}$$

Odhad společné variance $\sigma^2 = \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$
To je předpoklad testu.

Variabilita mezi skupinami (definovanými faktorem A): [among group mean square, group MS, MS_G]

$$SS_A = \sum_{i=1}^{k} n_i(\bar{X}_{i\bullet} - \bar{X}_{\bullet\bullet})^2$$

$$DF_A = k - 1$$

každé pozorování přispívá

$$MS_A = \frac{SS_A}{k-1} = \frac{\sum_{i=1}^{k} n_i(\bar{X}_{i\bullet} - \bar{X}_{\bullet\bullet})^2}{k-1}$$

Za platnosti H_0 je MS_A také odhadem společné variance σ^2.

Celková variabilita (už jen z tradice):

$$SS_{TOT} = \sum_{i=1}^{k}\sum_{t=1}^{n_i}(X_{it} - \bar{X}_{\bullet\bullet})^2 \qquad DF_{TOT} = N - 1 \qquad MS_{TOT} = \frac{SS_{TOT}}{DF_{TOT}} = \frac{SS_E + SS_A}{DF_E + DF_A}$$

Jednocestná analýza rozptylu

Říkáme, že celkovou variabilitu v datech rozkládáme na variabilitu vysvětlenou modelem (MS_A) + variabilitu reziduální (MS_E).

Platí, že MS_E je odhadem společné variability σ^2 (při dodržení předpokladu o shodnosti rozptylů).

Platí-li H_0 o shodnosti průměrů, potom je také MS_A odhadem σ^2, nezávislým na MS_E.

Testová statistika:

$$F = \frac{\frac{MS_A}{df_A}}{\frac{MS_E}{df_E}} \sim F_{df_A, df_E}$$

- Když H_0 platí a průměry jsou srovnatelné, potom je $F \doteq 1$.
- Když H_0 neplatí a alespoň jeden průměr je jiný než ostatní, potom je variabilita mezi průměry MS_A podstatně větší než uvnitř skupin a $F \gg 1$.

Jednocestná analýza rozptylu

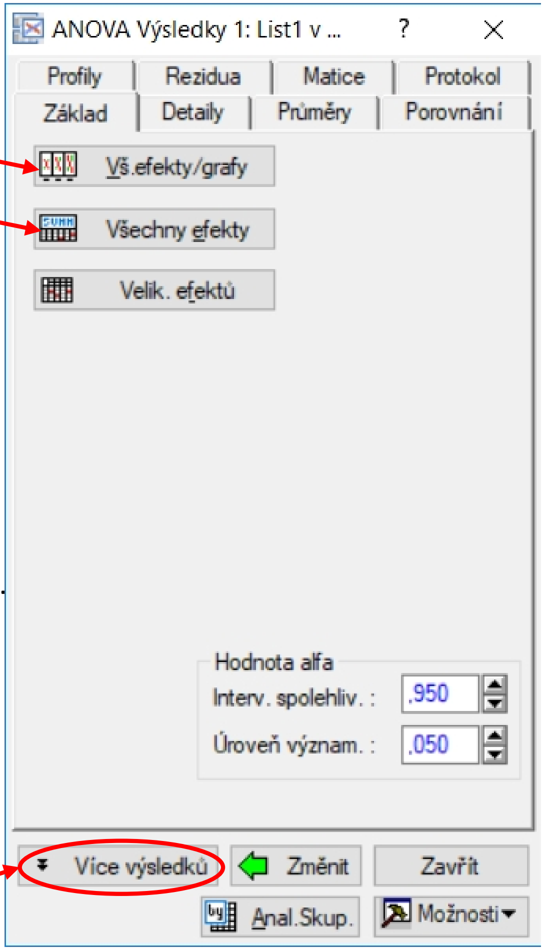
Příklad: koncentrace mědi v játrech ryb

STAT: *Statistiky* → *ANOVA* → *Jednofaktorová ANOVA* (+ *Select Cases*) → zadat proměnné → OK → *Všechny efekty*

součet čtverců   průměrný čtverec

Jednorozměrné testy významnosti pro konc (List1 v ANOVA °jatra)
Sigma-omezená parametrizace
Dekompozice efektivní hypotézy

Efekt	SČ	Stupně volnosti	PČ	F	p
Abs. člen	80,80625	1	80,80625	468,4327	0,000000
místo	3,32685	4	0,83171	4,8214	0,004008
Chyba	5,17510	30	0,17250		



Místo (jméno faktoru): test hypotézy, že $\mu_1 = \mu_2 = \ldots = \mu_k$.

Chyba [Error]: reziduální součet čtverců a odhad reziduálního rozp
Kontroluj počty stupňů volnosti.

Absolutní člen [Intercept]: test hypotézy, že celkový průměr = 0. V našem modelu nesmyslný test, pro prezentaci výsledků raději umazat.

Total: někdy také celkový součet čtverců a MS_{TOT}.

Jednocestná analýza rozptylu

Příklad: koncentrace mědi v játrech ryb

R:
```
> jatra.aov<-aov(konc~místo, data=jatra)
  summary(jatra.aov)
            Df Sum Sq Mean Sq F value  Pr(>F)
místo        4  3.327  0.8317   4.821 0.00401 **
Residuals   30  5.175  0.1725
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

univerzální funkce pro analýzu rozptylu

```
> oneway.test(konc~místo, data=jatra, var.equal=F)

        One-way analysis of means (not assuming equal variances)
data:  konc and místo
F = 6.0595, num df = 4.000, denom df = 14.419, p-value = 0.00451
```

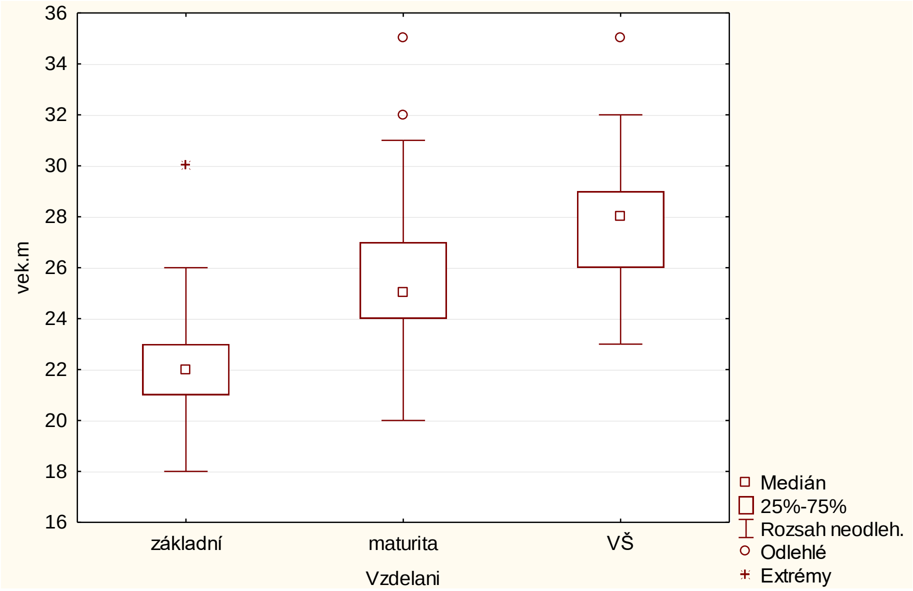
Jen analýza jednoho faktoru, možnost var.equal=F/T

```
> anova(lm(konc~místo, data=jatra))
Analysis of Variance Table
Response: konc
          Df Sum Sq Mean Sq F value   Pr(>F)
místo      4 3.3269 0.83171  4.8214 0.004008 **
Residuals 30 5.1751 0.17250
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

funkce porovnávající dva modely; model musím vytvořit předem

Ověření předpokladů analýzy rozptylu

1) **Vizuální ověření: boxploty, histogramy, q-q ploty po skupinách**
   - <u>Normalita</u> = hodnoty rozložené souměrně kolem průměru/mediánu, tedy mediány uprostřed krabice, oba vousy podobně dlouhé
   - <u>Shodnost rozptylů</u> = krabice odpovídající mezikvartilovému rozpětí/ směrodatné odchylce jsou podobně vysoké u všech skupin
   - Při velkém nesouladu zkusit <u>transformaci</u> (bude později) nebo <u>neparametrický test</u> nebo <u>zobecněné lineární modely</u> (např. počty něčeho málo početného /0 až 3 kusy/ budou mít spíš diskrétní Poissonovo rozdělení)



Ověření předpokladů analýzy rozptylu

2) Test shodnosti rozptylů ve skupinách = homoskedasticity

- **Leveneův test** – pracuje s absolutními odchylkami pozorování od skupinových průměrů $|X_{it} - \bar{X}_{i\bullet}|$ a na tyto hodnoty aplikuje analýzu rozptylu.

  H₀: průměrné odchylky pozorování od skupinových průměrů jsou stejné ve všech skupinách.

  Předpoklad: normální rozdělení hodnot v každé skupině.

- **Brown & Forsythe test** – modifikace Leveneova testu na mediány $\left|X_{it} - \tilde{X}_{i\bullet}\right|$. Takto je test robustnější vůči odchylce od normálního rozdělení dat.

- **Bartlettův test** – klasický test, jehož nevýhodou je značná citlivost na porušení předpokladu o normálním rozdělení hodnot ve skupinách.

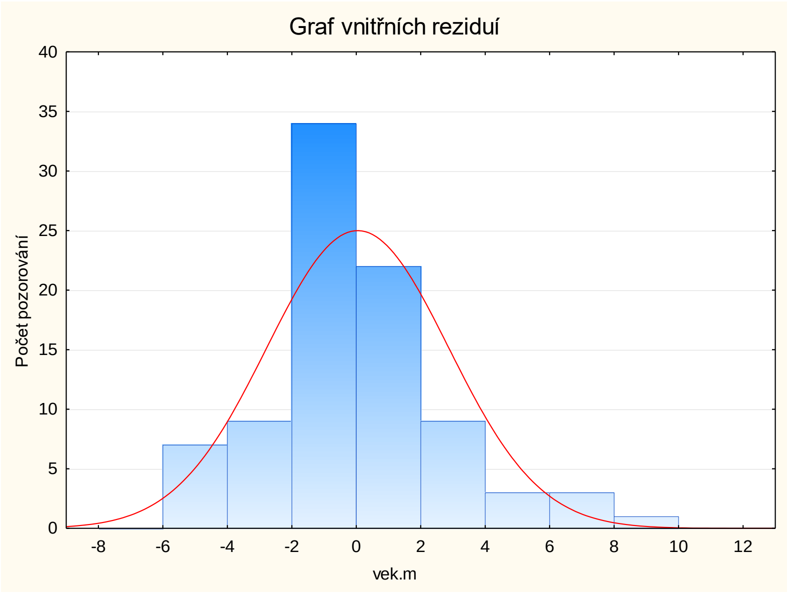
  Je založený na porovnání logaritmu reziduálního rozptylu s váženým průměrem logaritmů rozptylů jednotlivých skupin:

  $B = \frac{1}{C}\left((N-k)lnS^2 - \sum_{i=1}^{k}(n_i - 1)lnS_i^2\right)$, kde číslo ***C*** určují počty pozorování ve skupinách a vychází malinko větší než 1.

Ověření předpokladů analýzy rozptylu

3) Test normality – Shapiro-Wilkův test na reziduály

Není třeba testovat každou skupinu zvlášť, přestože předpoklady jsou právě takto formulovány. Jestliže jsme nezamítnuli hypotézu o stejných variancích (rozptylech) ve skupinách, znamená to, že skupiny se odlišují ve středních hodnotách (průměrech), ale pozorování jsou kolem průměru rozptýleny podobně. Vezmu proto všechny reziduály $|X_{it} - \bar{X}_{i\bullet}|$ z $\boldsymbol{k}$ skupin dohromady a studuji rozdělení například pomocí Shapirova-Wilkova testu. Tento postup je výhodný zejména v situaci, kdy jsou jednotlivé skupiny málo početné.



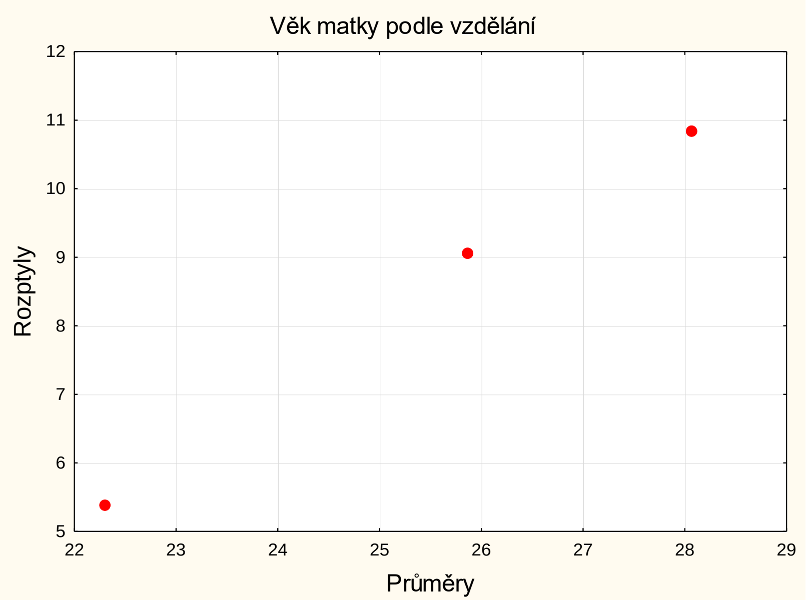
Ověření předpokladů analýzy rozptylu

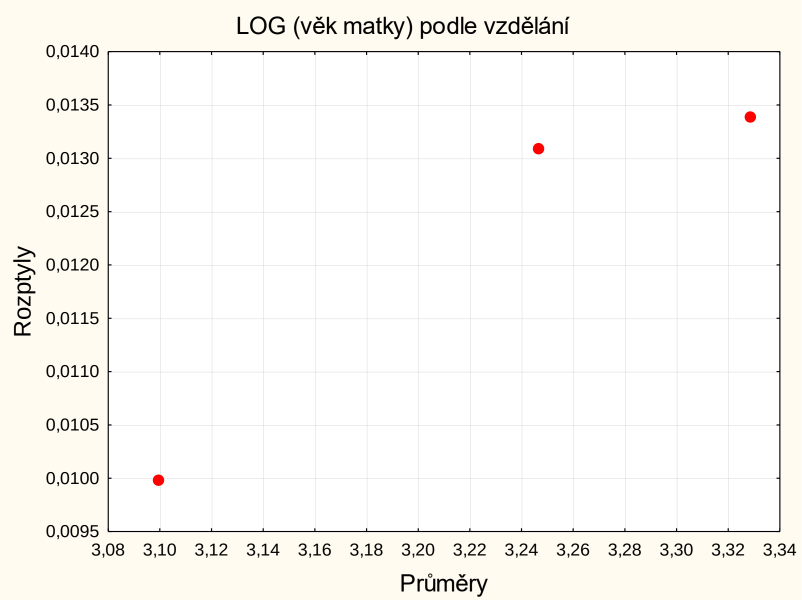
4) Kontrola vztahu rozptylu a střední hodnoty

Upozorní na možné problémy s nehomogenitou rozptylů. Zajímá nás situace, kdy s vyšším průměrem se zvětšuje také rozptyl. Toto je charakteristické zejména pro logaritmicko-normální rozdělení. Logaritmováním původních pozorování můžeme přiblížit rozptyly ve skupinách i celý výběr normálnímu rozdělení.

Častá vlastnost dat o koncentracích, hmotnosti a jiných objemových ukazatelů.

Vztah rozptylů a středních hodnot lze vypozorovat také z boxplotů.





Ověření předpokladů analýzy rozptylu

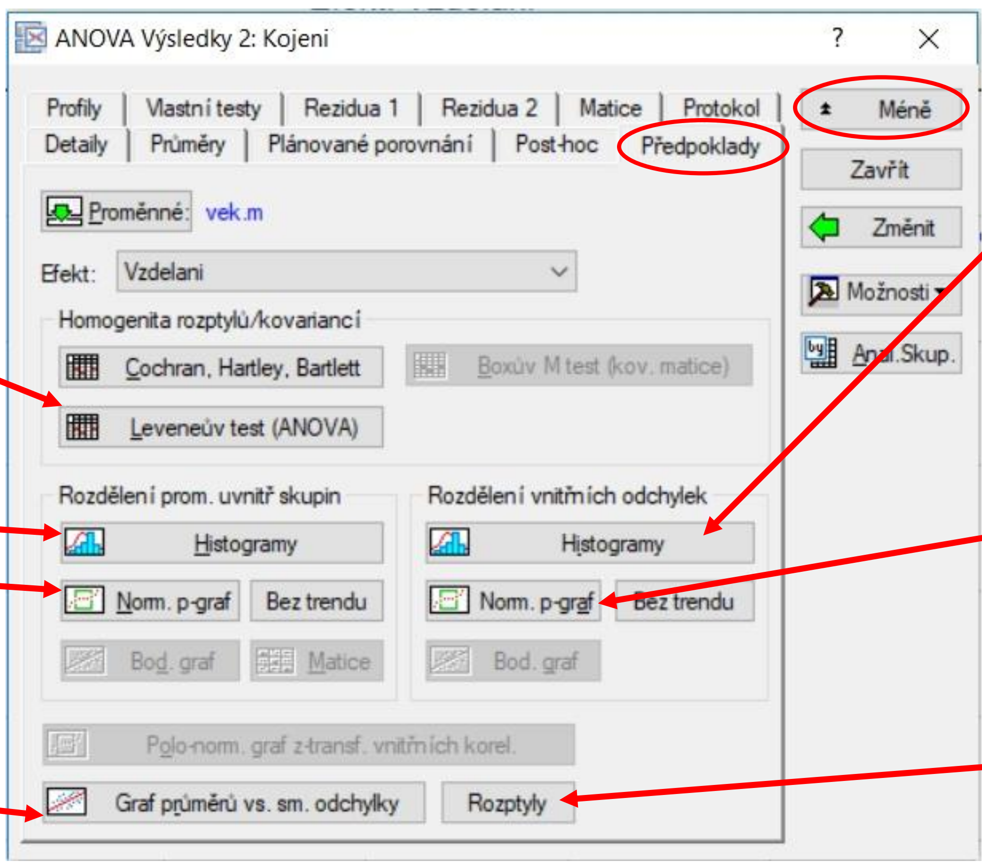
ANOVA je celkem robustní vůči narušení předpokladů, proto je jejich ověřování zlehčováno. Ověřování je však důležité například v těchto situacích:

- Mám jen málo pozorování v jednotlivých skupinách → dívám se hlavně na normalitu reziduálů a na vztah rozptylů a středních hodnot.
- Mám nevyvážený design [unbalanced design], tj. některé skupiny početné a jiné velmi nepočetné.
- Data mají výrazně šikmé rozdělení. Pomůže transformace?
- Leveneův F-test mohu posuzovat na hladině chyby 1. druhu 1 %.

Ověření předpokladů – možnosti softwaru

STATISTICA:

1. Zadám proměnné a spustím výpočet ANOVA
2. Zmáčknu tlačítko „Více výsledků“ (dole vlevo)
3. Zvolím záložku „Předpoklady“



Leveneův test

Histogramy a Q-Q ploty po skupinách

Vztah sm. odchylky a stř. hodnoty

Histogram reziduí (všechny skupiny dohromady). Bohužel bez Shapiro-Wilka.

Q-Q plot pro rezidua.

Vztah rozptylu a stř. hodnoty.

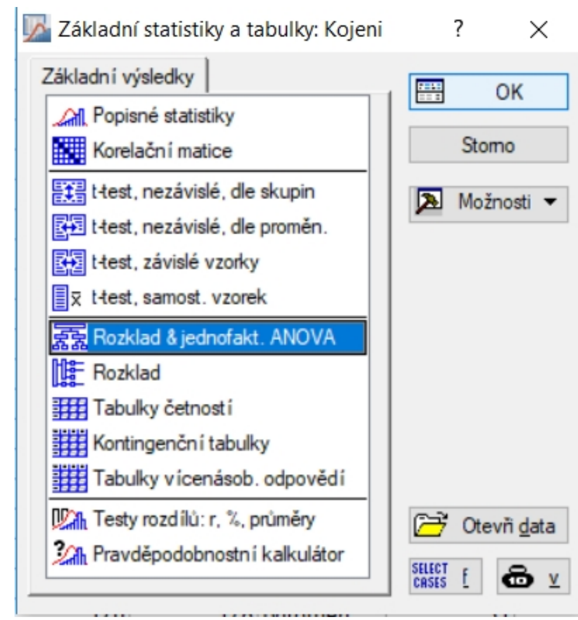
Ověření předpokladů – možnosti softwaru

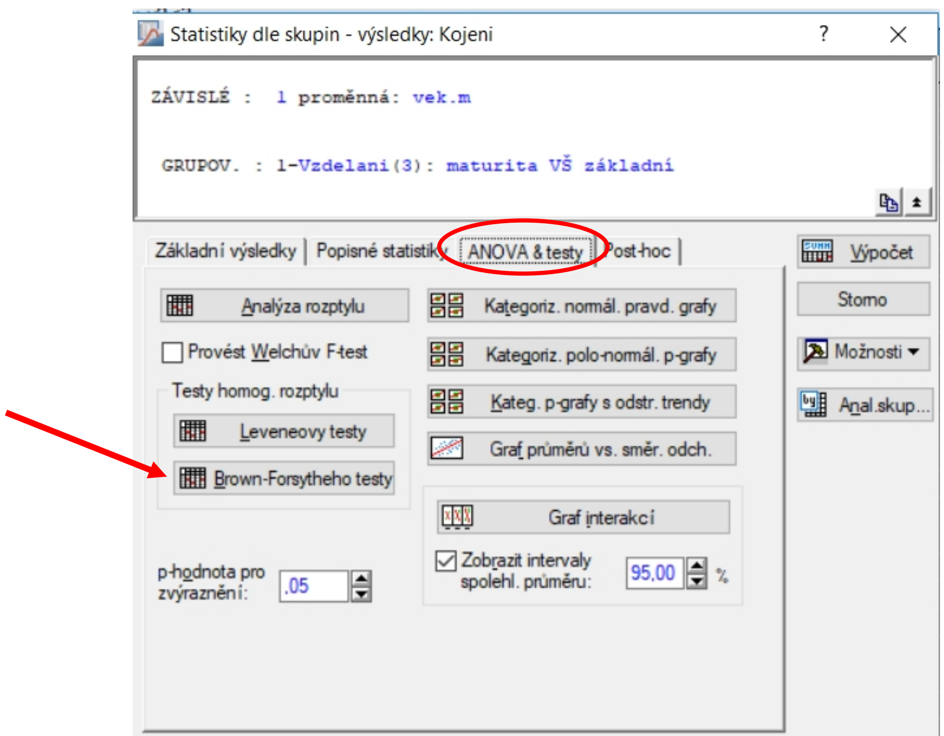
<u>STATISTICA</u>: Brownův-Forsytheův test homogenity rozptylů

Na jiném místě a pouze pro analýzu s jedním faktorem (jednocestná ANOVA):

Statistiky → Základní statistiky → Rozklad & jednofakt. ANOVA

Zadám proměnné a spustím výpočet (OK).





Ověření předpokladů – možnosti softwaru

R:

Mnohonásobná porovnání [multiple comparisons]

Pokud jsme v předchozím „celkovém“ testu zamítli nulovou hypotézu o shodnosti středních hodnot, zajímá nás, který z průměrů se liší od kterého.

Na tuto otázku použijeme takové testy, které mají ošetřeno nebezpečí vzrůstající chyby 1. druhu, tedy že označím za průkazný rozdíl mezi průměry i tam, kde ve skutečnosti rozdíl není (jen moje náhodné výběry náhodou pokrývají pravou a levou část spektra možných hodnot a proto se průměry zdají odlišné…)

Kolika testů se úvahy týkají?

Mám-li ***k*** skupin, potom

porovnávám $\frac{k\cdot(k-1)}{2}$ dvojic výběrů.

<u>Příklad</u>: ***k*** = 3 → 3*2/2 = 3

	Skup. 1	Skup. 2	Skup. 3
Skup. 1		1 vs. 2	1 vs. 3
Skup. 2			2 vs. 3
Skup. 3			

Mnohonásobná porovnání

Tukeyho metoda HSD

HSD = honest signifikance difference, tj. poctivě průkazný rozdíl

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

kde $q_{k,\,N\text{-}k}$ je kvantil studentizovaného rozpětí

- Odhad rozptylu je spočítán na základě variability všech skupin, nejenom těch dvou, které právě srovnáváme.
- Test byl původně navržen pro vyvážený design (všechny n_i stejná). Používáme také pro nestejná n_i , jen výsledek je potom spíše konzervativní, tzn. že skutečná pravděpodobnost chyby 1. druhu může být menší než zvolená hladina 5 %.
- Předpoklad homogenity variance je stále vyžadován a doporučení stejné velikosti skupin zde platí naléhavěji než pro hlavní test ANOVA modelu.

R: `TukeyHSD(x, which, ordered=FALSE, conf.level=0.95, ...)`

`x` není proměnná, ale objekt typu model, ve kterém jsou výsledky analýzy ANOVA `(aov)`

Mnohonásobná porovnání

Student-Newman-Keulsova metoda, SNK test

- Tento test seřadí průměry skupin podle velikosti a s využitím studentizovaného rozpětí určuje minimální rozdíl mezi průměry (s ohledem na počet pozorování), pro který už hypotézu o rovnosti průměrů zamítám.
- Tento test je sice silnější než Tukeyho metoda, bohužel má ale větší pravděpodobnost chyby prvního druhu (často větší než stanovenou α, např. 5 %)

Scheffého metoda je navržena pro obecné porovnávání, dokáže srovnat tedy nejen dvojice ale jakékoliv dvě skupiny úrovní proti sobě. Nicméně je méně citlivá než Tukeyho metoda, takže chceme-li porovnat pouze dvojice, je Tukeyho test lepší.

Duncanův test je založen na podobné myšlence jako SNK test, ale pravděpodobnost chyby prvního druhu je vztažena na jednotlivá porovnání. Proto pomocí tohoto testu získáme nejvíce průkazných rozdílů, ale nedodržíme pravděpodobnost chyby prvního druhu.

Mnohonásobná porovnání – prezentace výsledků

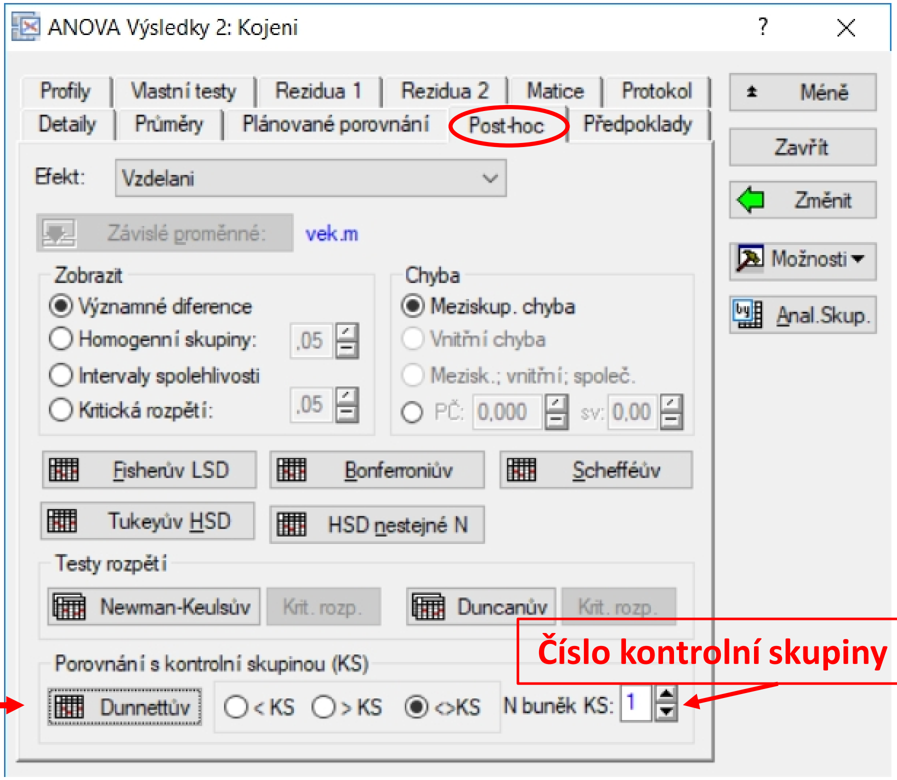
L&S 163

- Situace, kdy skupina se s „prostředním“ průměrem neliší ani od menšího ani od většího průměru, zatímco „malý“ a „velký“ průměr se odlišují průkazně. Není ale možné, aby „prostřední“ skupina byla totožná s „menší“ i „větší“ skupinou. Došlo zde zřejmě k chybě druhého druhu (nezamítám hypotézu, která neplatí). Je to dáno nejistotou způsobenou nedostatkem pozorování.
- Stává se, že hlavní F-test ANOVA modelu zamítá hypotézu o rovnosti průměrů, ale při mnohonásobných porovnáních nedostanu žádný průkazný rozdíl. To je dáno faktem, že mnohonásobné testování nemá takovou sílu jako hlavní F-test. Dá se ale očekávat, že data s větším počtem pozorování by pravděpodobně odhalila průkazné rozdíly i v mnohonásobných porovnáních.

Mnohonásobná porovnání

Dunnettův test

- Užitečný pro situaci, kdy mám skupinu kontrolní a skupiny „se zásahem“.
- Je to varianta Tukeyho testu, testová statistika je stejná. Síla Dunnettova testu je ale větší než u Tukeyho testu, protože porovnávám méně dvojic.
- Protože kontrola vstupuje do testu vícekrát, je třeba, aby byla odhadnuta přesněji než „zásahové“ skupiny. Doporučuje se proto, aby počet pozorování v kontrole byl o něco méně než $\sqrt{k-1}$krát větší než v jiné skupině a ostatní skupiny se nelišily ve své velikosti. ***k*** je počet skupin včetně kontroly.



Volba Dunnettova testu s možností volby jednostranných testů

Mnohonásobná porovnání

Bonferroniho korekce

Vynásobí výslednou p-hodnotu počtem testů k*(k-1)/2 a pak teprve porovnávám s pravděpodobností $\boldsymbol{\alpha}$ (chyby 1. druhu).

Bonferroniho test

Klasický test, který upravuje dvouvýběrový t-test takto:

Hypotézu o rovnosti průměrů $\bar{X}_{i\bullet}$ a $\bar{X}_{j\bullet}$ zamítám, když

$$\left|\bar{X}_{i\bullet} - \bar{X}_{j\bullet}\right| \geq t_{N-k}\left(1 - \frac{\alpha}{r}\right) \cdot \sqrt{S_E^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

$\boldsymbol{r}$ je počet všech porovnávaných dvojic, tedy r = k*(k-1)/2.

Odhad rozptylu $\boldsymbol{S_E^2}$ je odvozen z variability všech pozorování.

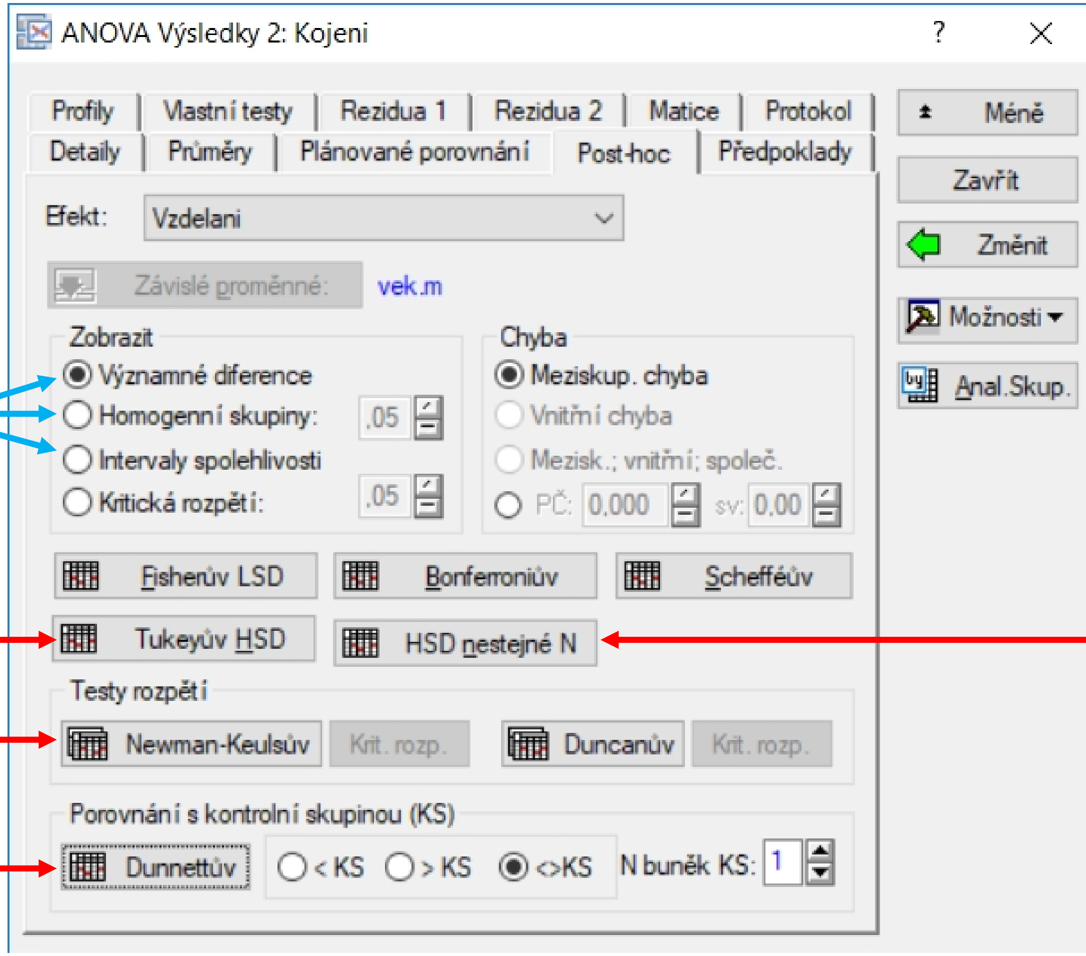
Fisherův LSD test

Odpovídá sérii t-testů a pravděpodobnosti chyb prvního druhu se sčítají.
Nedoporučený test.

Mnohonásobné porovnávání – zadání v softwaru

STAT:

Výsledkové tabulky
v různém formátu



R: `TukeyHSD()`

Mnohonásobné porovnávání – poznámky k chybě 1. druhu

Komentář k problematice porovnávání *a posteriori* (skupiny porovnávám až po provedení celkového F-testu) a *a priori* (předem vyberu skupiny, které mě zajímají) najdete v učebnici Biostatistika, Lepš & Šmilauer (2016) na straně 160. Autoři odkazují dále na Biometry, Sokal & Rohlf (2010, 4. vydání) str. 228-237.

Dále tam najdete komentář ke skutečné pravděpodobnosti chyby 1. druhu v situaci, kdy provádím vícero testů na jedněch datech (stále strana 160). *Comparisonwise Type I error rate, experimentwise Type I error rate, familywise Type I error rate.*

Neparametricky: Kruskalův-Wallisův test

Test založený na pořadí [analysis of variance by ranks].

- Předpoklady: výběry jsou vzájemně i „uvnitř“ nezávislé, tedy zcela náhodné experimentální uspořádání. Data nejlépe spojitá, ale lze testovat i pro data na ordinální stupnici.
- Pokud porovnávám jen 2 výběry, budou se čísla lišit od výsledků Mann-Whitneyova testu.
- Nulová hypotéza: distribuční funkce všech skupin jsou stejné, tedy speciálně také všechny mediány jsou stejné.
- Testová statistika:

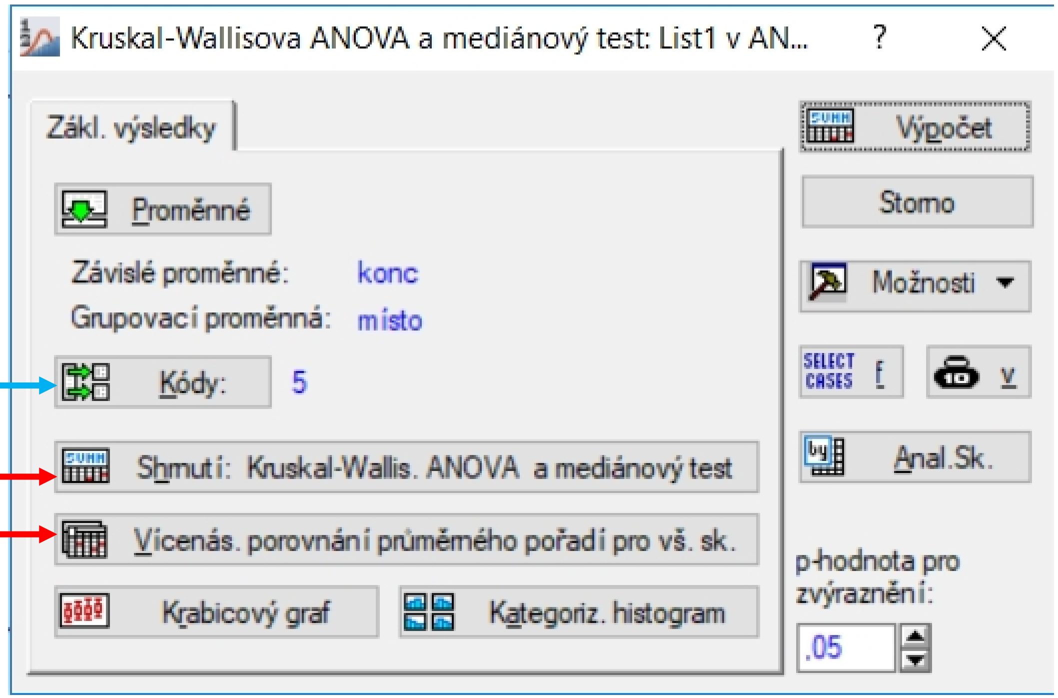
$$\boldsymbol{H = (N-1)\frac{\sum_{i=1}^{k} n_i (R_{i\bullet} - R_{\bullet\bullet})^2}{\sum_{i=1}^{k}\sum_{t=1}^{n_i}(R_{it} - R_{\bullet\bullet})^2} \sim \chi^2_{(k-1)}}$$

Neparametricky: Kruskalův-Wallisův test

STAT:

Statistiky → Neparametrické statistiky → Porovnání více nezávislých vzorků (skup.)

Tady snadno změním pořadí úrovní v boxplotu



R: **`kruskal.test`**`(formula, data, subset, na.action, ...)`

Neparametricky: mnohonásobná porovnání

STATISTICA: metoda podle knihy Siegel, S., & Castellan, N. J. (1988). Nonparametric statistics for the behavioral sciences (2nd ed.) New York: McGraw-Hill.

Testová statistika:

$$Z_{ij} = \frac{|R_{i\bullet} - R_{j\bullet}|}{\sqrt{\frac{N\,(N+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \sim N(0,1)$$

- Nezohledňuje množství shodných pořadí.
- Úpravu na mnohonásobné srovnání provede pomocí Bonferroniho korekce: ***p***-hodnota = ***p***-hodnota **** k*(k-1)/2***, tedy krát počet testovaných dvojic

R: v knihovně **`pgirmess`** funkce

```
kruskalmc(x ~ groups, data= , probs=0.05)
```

Další knihovna: **`multcomp`**