

# Zadání Cvičení #7

**Důležité!** V dnešním cvičení naleznete opět na Moodle ke stažení dva data sety, [data.csv](#) a [large\\_dataset.csv](#). **Zopakujte si princip vyšetření potřebné velikosti vzorku, který jste se naučili minule.**

**Popis dat:** Data jsou získána od dvou skupin:

- Pacienti s [Huntingtonovou nemocí](#) (HN), kteří **jsou** léčeni pomocí antipsychotik<sup>1</sup>, label **HD\_AP**
- Pacienti s Huntingtonovou nemocí, kteří **nejsou** léčeni antipsychotikou, label **HD\_NAP**

Obě skupiny podstoupily vyšetření řeči pomocí volného monologu. Řečové nahrávky byly analyzovány a ze záznamu byly vypočteny hodnoty parametru **RST** (*Rate of Speech Timing*, jednotka  $s^{-1}$ ), který popisuje rychlost řeči pomocí identifikace znělých (voiced), neznělých (unvoiced) a přerušených (pause) segmentů v záznamu. Detailní popis parametru RST a další informace viz [Hlavnička et al. 2017](#) (Reference).

Pro každého pacienta máme k dispozici dva *kategorické* údaje:

- Sloupec **group** (příslušnost ke skupině): HD\_AP nebo HD\_NAP (popis výše).
- Sloupec **gender** (pohlaví): F (žena, female) nebo M (muž, male).

Dále máme k dispozici údaje:

- Sloupec **medication**: Množství antipsychotických léků, které pacienti užívají<sup>2</sup> (jednotka miligramy, *mg*).
- Sloupec **motor**: Hodnocení stavu motoriky na škále UHDRS<sup>3</sup>.
- Sloupec **dysarthria**: Hodnocení severity dysartrie na škále UHDRS.

Zadání úlohy	body
<p>Použijte program <b>GPower</b>, abyste zjistili, jak velký vzorek je potřeba pro vaši analýzu. Předpokládejte, že budete testovat data pomocí dvoucestné ANOVY.</p> <p>Tentokrát nemáme k dispozici data u přechozích studií pro odhad populačních statistik, proto cílovou sílu efektu zvolíme jako „velkou“. <b>Čemu numericky odpovídá „velká“ konzultujte s manuálem programu GPower.</b> Sílu efektu budeme chtít dosáhnout pro hladiny <math>\alpha = 0.05</math> a <math>\beta = 0.2</math>.</p> <p><b>Rozhodněte se, zdali použijete malý (<math>n = 22</math>) nebo velký (<math>n = 38</math>) dataset.</b></p> <p>Pokud by ani jeden dataset svojí velikostí nedosahoval na požadovanou vypočtenou velikost vzorku, proveďte všechny další analýzy stejně, ale k této části <b>slovně napište, jak to ovlivní interpretaci vašich výsledků s ohledem na chybu II. typu.</b></p>	<b>1</b>

<sup>1</sup> **Antipsychotika** = třída léků určená k léčbě psychóz (bludů, halucinací, paranoie, ...).

<sup>2</sup> *Chlorpromazine equivalent*, tzv. celková síla medikace která by odpovídala užívání x miligramů typického antipsychotického léku [chlorpromazinu](#).

<sup>3</sup> **UHDRS** = *Unified Huntington's Disease Rating Scale*, škála pro hodnocení severity Huntingtonovy choroby.

Zadání úlohy	body
<p>V této úloze je vizualizace dat zcela dobrovolná a nebude hodnocena, nicméně ji samozřejmě zcela doporučujeme zejména k lepšímu porozumění pozorovaných efektů.</p>	0
<p>Pokuste se odpovědět pomocí statistického testování na otázky níže. Zajímá vás rozlišitelnost dat závislé proměnné pomocí dostupných faktorů. Pro statistické testování využijte funkci MATLABu <code>anovan</code>, případně <code>anova</code>.</p> <p><b>Na následující dvě otázky odpovězte pomocí jediného testu.</b></p> <ol style="list-style-type: none"> <li>1. <i>Mají pacienti s HN beroucí antipsychotickou medikaci statisticky významně změněnou rychlost řeči oproti pacientům, kteří antipsychotika neberou?</i></li> <li>2. <i>Lze pozorovat statisticky významné změny v rychlosti řeči způsobené antipsychotickými léky specificky (pouze) pro jedno z pohlaví?</i></li> </ol> <p><b>Následující otázku musíte řešit testem zvlášť od prvních dvou.</b></p> <ol style="list-style-type: none"> <li>3. <i>Pokud bychom brali hodnocení stavu motoriky na škále UHDRS jako nezávislý kovariát (spojitý faktor), jaký bude mít vliv na výsledky z první otázky?</i></li> </ol> <p><b>U otázek v této i následující části si dejte záležet na korektním provedení testu a interpretaci výsledků.</b></p> <ul style="list-style-type: none"> <li>• Testování normality/homoskedasticity vám dá informaci o tom, jak budou výsledky ovlivněné chybou I. typu, pokud vstupní data nebudou normální, protože zde není možnost použít neparametrickou metodu pro více-faktorové testování.</li> <li>• Výsledky, jako vždy, korektně reportujte.</li> </ul>	2
<p>Rozdělme si pacienty na ty s postiženou řečí a ty bez postižené řeči – to můžeme provést na základě klinického hodnocení řeči skrytého pod položkou <b>dysarthria</b>, kde hodnota 0 znamená nepostiženou řeč, všechny ostatní hodnoty pak řeč (různě) postiženou.</p> <p><i>Mají pacienti s poruchou řeči ve srovnání s pacienty bez poruchy řeči významně zhoršené hodnoty motorického hodnocení UHDRS, pokud budeme brát v potaz i vliv toho, jestli berou antipsychotika, nebo ne?</i></p>	1

## Teorie:

V tomto cvičení si vyzkoušíte práci s vícecestnou ANOVA (*N-way ANOVA*, nebo také *Factorial ANOVA*). Ta se používá v případech, kdy máme:

- 1 závislou proměnnou (*dependent variable*, zde například naměřené hodnoty parametru RST)
- 2 a více **faktorů**, nezávislých proměnných, např.:
  - Typ pacienta (*group*): léčení/neléčení
  - Pohlaví (*gender*): muž/žena

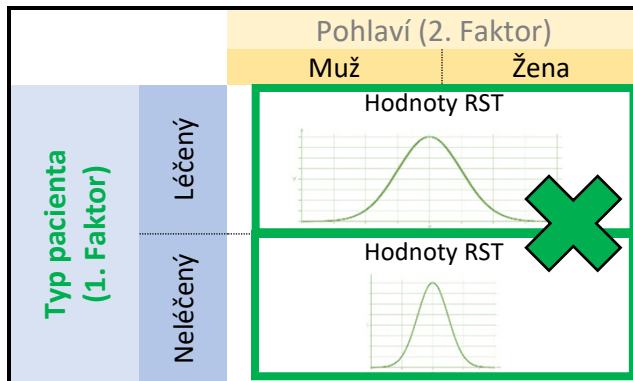
Faktory dělí naše data na **bloky** – skupiny, jejichž počet je určen součinem „velikostí“ jednotlivých faktorů, tedy např. zde  $2 \times 2 = 4$  kombinace: léčení-muž, léčení-žena, neléčení-muž, neléčení-žena. Pokud máme **dva faktory**, můžeme také používat název **dvoucestná ANOVA** (*2-way ANOVA*) apod.

Výhodou vícecestné ANOVA je možnost vyšetřit několik hypotéz jedním testem a vyšetřit interakce mezi jednotlivými faktory. **Interakce mezi faktory** (značeno např. *GROUP\*GENDER*, pro interakci mezi skupinami a pohlavím) znamená, že existuje signifikantní rozlišitelnost na úrovni rozdělení pomocí obou faktorů, jak ilustrují tabulky na další straně. **V praxi to znamená, že když například 1. faktor nevykáže signifikantní rozdíl, ale interakce ano, je možnost, že skupiny léčený/neléčený lze pomocí RST rozlišit, ale jen pokud budeme zkoumat muže nebo ženy separátně.**

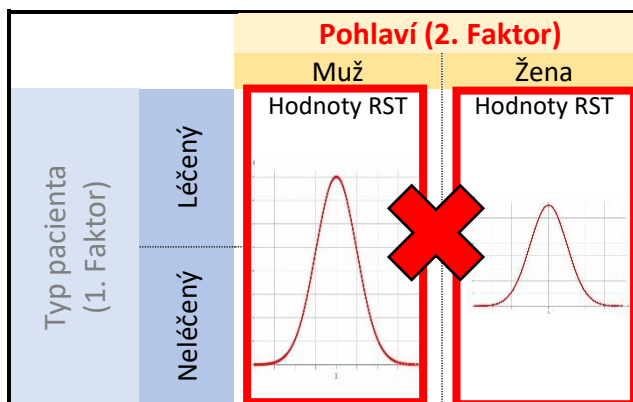
## Reference

Hlavnička, J., Čmejla, R., Tykalová, T., Šonka, K., Růžička, E., and Ruzs, J. (2017). *Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder*. Scientific reports, 7, 12, <https://doi.org/10.1038/s41598-017-00047-5>.

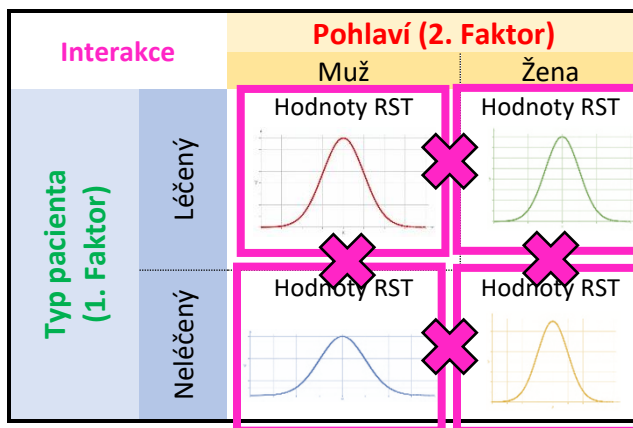
## Dvoucestná (2-way) ANOVA: druhy porovnávání



1. Rozdíly mezi bloky vytvořenými rozdělením podle prvního faktoru: typu skupiny. Uvažují se pouze rozdíly v RST mezi daty seskupenými do skupiny léčených a skupiny neléčených pacientů.



2. Rozdíly mezi bloky vytvořenými rozdělením podle druhého faktoru: pohlaví. Uvažují se pouze rozdíly v RST mezi daty seskupenými do skupiny mužů a skupiny žen.



3. Rozdíly mezi bloky vytvořenými rozdělením podle obou faktorů zároveň. Uvažují se rozdíly mezi jakoukoli dvojicí z bloků muž-léčený, muž-neléčený, žena-léčená, žena-neléčená. Pokud vyjde interakce jako statisticky signifikantní, dává nám to informaci, že existuje významný rozdíl mezi alespoň jednou dvojicí, ale neřekne nám mezi kterou.