

Zadání Cvičení #6

Důležité! V dnešním cvičení naleznete na Moodle ke stažení dva data sety, [data.csv](#) a [large_dataset.csv](#). Předtím, než přistoupíte k načítání dat do Matlabu (klasicky pomocí funkce `readtable`, viz předešlá cvičení), musíte se rozhodnout, který z těchto dvou datasetů použijete.

Toto rozhodnutí ilustruje proces přípravy výzkumného experimentu, kdy je v praxi získávání (nových) dat jedna z nejnáročnějších úloh ať už v medicíně, kdy je nutné svolávat pacienty na schůzky, organizovat pracovníky pro sběr dat, zajistit prostory atd. – tak i v jiných oborech, kde je nutné zajistit přístup k měřené věci, ať už je to stroj nebo fyzická lokace, doprava na místo, správné nastavení podmínek pro měření atd.

Proto je dobré před začátkem experimentu vyhodnotit, kolik dat **minimálně** vůbec potřebujete, aby byly vaše případné výsledky relevantní. Vyjdete přitom z výzkumů, které se zabývaly stejným tématem, nebo se pokusíte potřebné populační parametry odhadnout.

Popis dat: Data jsou získána od dvou skupin:

- Pacienti s Parkinsonismem¹ způsobeným otravou manganem po užívání efedronu² (label **EP**)
- Kontrolní skupina zdravých lidí (label **HC**)

Obě skupiny podstoupily vyšetření řeči:

1. Pomocí *diadochokinetického* (DDK) testu – na nahrávkách byly vyhodnoceny tyto parametry: ([Novotný et al. 2014](#), viz Reference)
 - **DDKG** – *DDK regularity*, pravidelnost sledu hlásek *pa-ta-ka*, bezrozměrná jednotka (-)
 - **DDKR** – *DDK rate*, rychlost opakování hlásek *pa-ta-ka*, jednotka *syll/s* (hlásky za sekundu)
2. Pomocí testu prodloužené fonace hlásky “a” – na nahrávkách byly vyhodnoceny parametry: (měřeno pomocí programu *PRAAT*, [Boersma and Weenink 2017](#), viz Reference)
 - **stdF0** – směrodatná odchylka fundamentální frekvence (značeno také **F0 SD**), jednotka *semitones* (půltóny)
 - **jitter** – odchylka od pravidelné periodicity chvění hlasivek při fonaci, jednotka %

Statistické charakteristiky pro všechny výše uvedené parametry **najdete v tabulce IV** ve studii [Rusz et al. 2011](#), která je **dostupná ke stažení ve složce cvičení** na Moodle. (budete jí potřebovat!)

Všem pacientům ze skupiny EP byl ještě ohodnocen stav jejich motoriky na škále NNIPPS³ a v tabulce najdete hodnoty pro položky Speech (řeč), [Bradykinesia](#), [Dystonia](#) a [Pyramidal](#).

¹ [Parkinsonismus](#) = syndrom, soubor motorických postižení (klidový třes, zpomalení pohybů, ztuhlost a posturální labilita) typických pro Parkinsonovu nemoc ale vyskytující se u pacientů i z celé řady jiných příčin.

² [Efedron](#) = droga podobná pervitinu, působí stimulační a euforické účinky. Náhražka kokainu.

³ NNIPPS = *The Natural History and Neuroprotection in Parkinson Plus Syndromes scale*

Zadání úlohy	body
<p>Pro tuto část budete potřebovat program GPower, který je dostupný volně ke stažení zde v sekci <i>Download</i>. Na stránkách je také dostupný ke stažení anglický manuál.</p> <p>Pro všechny 4 akustické parametry (DDKG, DDKR, stdF0 a jitter) určete minimální celkovou velikost vzorku (<i>sample size</i>) pro zodpovězení otázek z předposlední části zadání. Použijte informace z předchozí studie Rusz et al. 2011, jejíž text je dostupný ke stažení ve složce cvičení na Moodle.</p> <ul style="list-style-type: none"> • Pro jitter použijte statistické hodnoty parametru jitter:local z tabulky • <u>Nezaměňte stdF0 měřené z fonačního testu za F0 SD z jiného testu!</u> • Nespleťte si sloupečky pro <i>mean</i> a <i>std</i> obou skupin <p>Pro výpočet minimální celkové velikosti vzorku použijte program GPower: (podrobný návod najdete na konci tohoto PDF)</p> <ul style="list-style-type: none"> • Rozmyslete si, jaké statistické testy využijete na zodpovězení otázky č. 1 z předposlední části tohoto cvičení. • Při výběru testů zohledněte normalitu parametrů pro obě skupiny a podle toho vyberte parametrický nebo neparametrický test. Když ale budete v GPower počítat potřebný sample size, vypočtete ho jen pro parametrický test – výsledek se oproti tomu neparametrickému nebude téměř lišit. Toto se ale týká jen výpočtu sample size v GPower! • Použijte hodnoty: $\alpha = 0.05$, $\beta = 0.2$ (tedy $1-\beta = 0.8$) <p>Podle výsledné potřebné minimální velikosti vzorku se rozhodněte, které parametry má smysl analyzovat s daty, která máte k dispozici (ať už z velkého nebo malého datasetu). K parametrům, na které budou poskytnutá data stačit, vyberte co nejmenší dataset, tak, jak vám potřebná velikost vzorku dovolí.</p> <p>Když budete mít data o pár jednotek menší, než vám vyjde minimální velikost vzorku, nemusíte si s tím dělat starosti.</p> <p>Do Moodle запиšte výsledné vypočtené minimální potřebné velikosti vzorku, a jaké datasety jste se rozhodli použít s jakými parametry, případně jaké parametry jste se rozhodli vyřadit.</p>	1
<p>Vybraná data vhodným způsobem vizualizujte. Nejlépe využijte KDE nebo normální PDF, tak, aby byly zřetelné případné rozdíly mezi skupinami.</p> <p>Do Moodle odevzdejte obrázek s vizualizací a slovně zhodnoťte, zdali je mezi skupinami vidět nějaký efekt skupin (size effect).</p> <p>Kvantifikujte efekt skupin pomocí Cohenova <i>d</i> (využijte svojí naprogramovanou funkci z předchozích cvičení) a slovně do Moodle uveďte, jak velký tento efekt je.</p>	1

<p>K následujícím otázkám navrhnete hypotézy, vyberte vhodné testy, nastavte hladinu statistické významnosti korigovanou pro ošetření chyby I. typu <u>a korektně</u> reportujte svoje výsledky do Moodle.</p> <p>Mějte prosím na paměti, že zde budete provádět několik hypotéz/testů, které budou odpovídat na JEDNU otázku (jako kdyby vás zajímalo, jestli nějaká barva žužu fazole způsobuje akné a prováděli byste test pro každou barvu).</p> <ol style="list-style-type: none"> 1. Pokud kvalitu řeči popíšeme pomocí naměřených hodnot DDKG, DDKR, stdFO a jitter, můžeme najít statisticky významný rozdíl mezi kvalitou řeči zdravých lidí (ve skupině HC) a kvalitou řeči lidí s Parkinsonismem po otravě efedronem (ve skupině EP)? 2. Vyberte si jednu položku z dotazníku NNIPPS. Ověřte, zda u členů skupiny EP existuje statisticky významný lineární vztah mezi hodnotami vybrané položky z dotazníku a kvalitou řeči, popsanou pomocí naměřených hodnot DDKG, DDKR, stdFO a jitter. <ul style="list-style-type: none"> o Pro analýzu použijte stejný dataset, jaký jste vybrali pro zodpovězení otázky z předchozího bodu. 	1.5
<p>Připomenutí: Všechny výsledky testů запиšte v korektní formě, včetně názvu použitého testu, případně jeho parametrů (jako např. hladina statistické významnosti). Pokud nevíte, jak správně reportovat výsledek nějakého testu, stačí to jednoduše vygooglit. Pokud máte na výběr a nevíte který styl formátování použít, použijte styl APA (American Psychological Association).</p>	
<p>Zjistěte pomocí programu GPower dosaženou sílu (<i>Power</i>, $1-\beta$) u testů, které jste provedli pro zodpovězení otázky č.2 v předchozím bodě.</p> <p>Využijte pro výpočet korigovanou hladinu statistické významnosti (pokud jste jí korigovali), celkovou velikost použitého vzorku (záleží na použitém datasetu) a vypočtené hodnoty.</p> <p>Nápověda: Typ analýzy = post-hoc.</p>	0.5

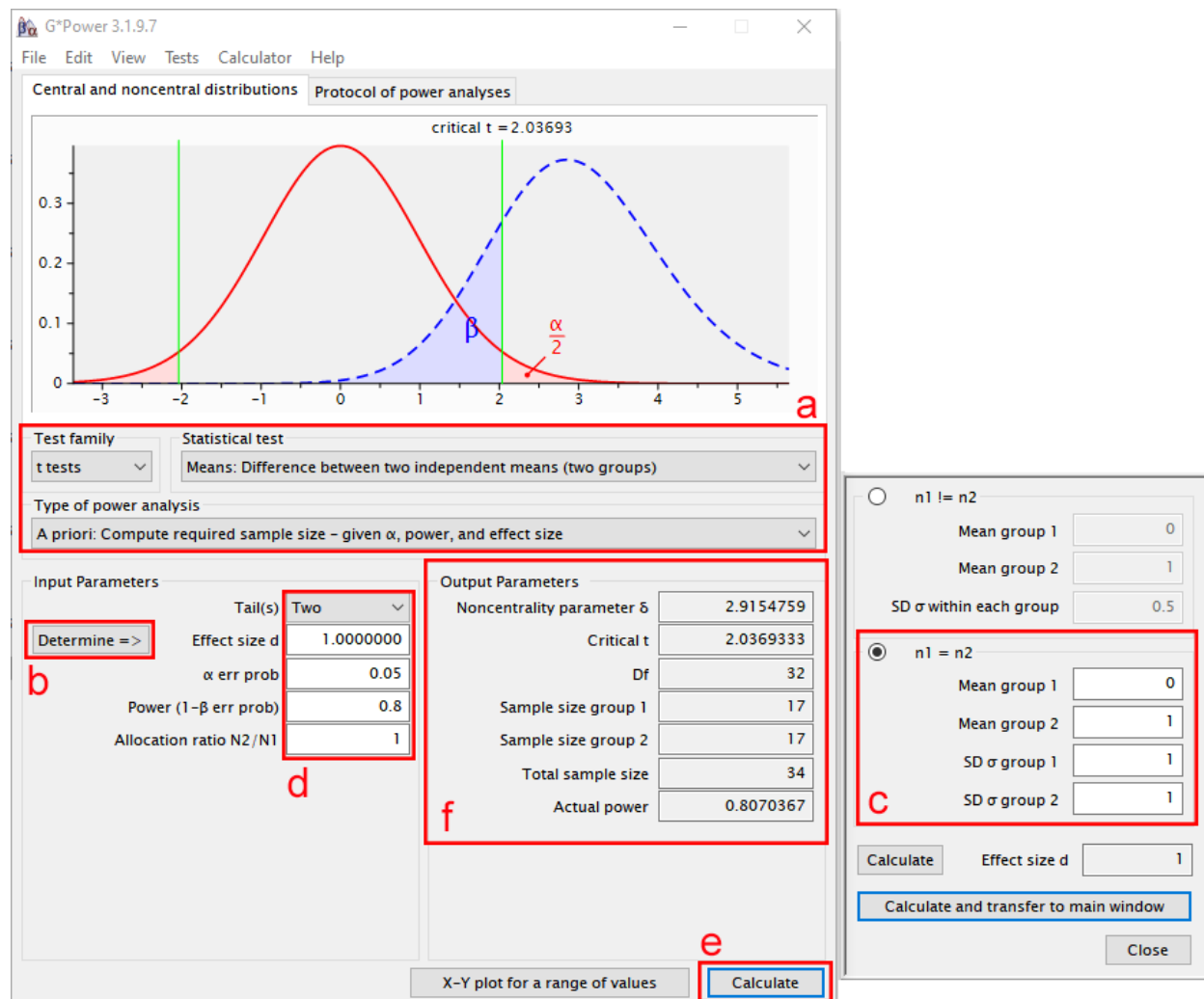
Reference

Boersma, P. and Weenink, D. (2017). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.30, retrieved 22 July 2017 from <http://www.praat.org/>

Novotný, M., Ruz, J., Čmejla, R., and Růžička, E. (2014). *Automatic evaluation of articulatory disorders in Parkinson's disease*. IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 22, 1366-1378, DOI: [10.1109/TASLP.2014.2329734](https://doi.org/10.1109/TASLP.2014.2329734)

Ruz, J., Čmejla, R., Růžičková, H., and Růžička, E. (2011). *Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease*. The journal of the Acoustical Society of America, 129, 350-367, DOI: [10.1121/1.3514381](https://doi.org/10.1121/1.3514381)

Nastavení v programu GPower



a) Základní nastavení:

1. Test Family: Výběr rodiny testů (t tests = t-testy, Wilcoxon testy, F tests = ANOVA apod.)
2. Statistical test: Výběr konkrétního testu z rodiny (normální/nenormální, párový/nepárový...)
3. Type of power analysis: Výběr typu analýzy

b) Tlačítko pro zobrazení okna vpravo, kde lze zadat hodnoty pro výpočet síly efektu (*effect size*)c) Výpočet *effect size*: Zadejte odhad statistických parametrů populací – použijte data z předchozích studiíd) Zadejte další parametry výpočtu, jedno/obou-strannost testu, hodnoty α a $(1-\beta)$ a poměr velikostí vzorků obou skupin ($N2/N1$). Effect size se doplní po výpočtu z bodu c).

e) Zmáčkněte pro výpočet

f) Výsledek analýzy: zajímá vás zejména „total sample size“, tedy celková velikost vzorku (součet velikostí vzorků obou skupin)