

Zadání Cvičení #8

Popis dat: Pracovní data jsou uložena v souboru **data.csv**, který je k dispozici ke stažení na Moodle stránce tohoto předmětu, ve složce příslušného cvičení. Pro načtení dat do Matlabu využijte funkci `readtable`.

Data jsou ve formátu tabulky, která obsahuje data od pacientů s Parkinsonovou nemocí (label **PD**). Tabulka obsahuje identifikační kód pacienta, pohlaví, identifikátor skupiny a řadu parametrů:

- Parametry vyhodnocené počítačem z nahrávek řeči během diadochokinetického (DDK) testu, viz Reference [Novotný et al. 2014](#):
 - **DDKR – DDK rate**: rychlost opakování hlásek *pa-ta-ka*, jednotka *syll/s* (hlásky za sekundu).
 - **VOT – Voice Onset Time**: parametr popisující motorickou kontrolu nad délkami souhlásek, jednotka milisekundy *ms*.
- Parametry vyhodnocené pomocí programu PRAAT (Reference [Boersma and Weenink 2017](#))
 - Z nahrávek souvislé řeči:
 - **stdPWR** – Směrodatná odchylka intenzity řeči, jednotka dB.
 - **stdF0** – Směrodatná odchylka F0, jednotka půltóny – *semitones*.
 - Z nahrávek testu prodloužené fonace hlásky “a”:
 - **jitter** – odchylka od pravidelné periodicity chvění hlasivek při fonaci, jednotka %.
- **PPT – Purdue Pegboard Test**: manuální test, při kterém vyšetřovaný vkládá či vytahuje kolíčky do speciální desky. Hodnotí se tak zručnost, jemná i hrubá motorika, koordinace, bez jednotky.
- **AmpDec – Amplitude Decrement**: hodnocení změny maximální vzdálenosti mezi prsty během *finger tapping* úlohy, která byla zaznamenána 3D kamerovým systémem. Vyhodnocené pomocí počítače, detaily k výpočtu parametru viz Reference [Krupička et al. 2020](#), bez jednotky.
- **VFT – Video Finger Tapping**: klinické expertní hodnocení *finger tapping* úlohy z videozáznamu (bez jednotky).
- Klinické hodnocení neurologem prostřednictvím škály UPDRS III (bez jednotky):
 - **UPDRS III**: celkové skóre z dotazníku.
 - **UPDRS III axial**: souhrnné skóre z položek, které hodnotí axiální¹ motoriku.
 - **UPDRS III bradykinesia**: souhrnné skóre z položek, které hodnotí *bradykinezi*².
- **L-dopa equivalent**: Síla medikace - množství [Levodopy](#), které by mělo stejný výsledný efekt jako kolektivní účinek všech léků, které pacient bere za jeden den, jednotka miligramy, mg.

¹ **Axiální** – týkající se těch částí těla, které se nachází na svislé ose (hlava, krk, hrudník, trup, ...), opak je apendikulární (týkající se končetin)

² **Bradykinezie** – zpomalení pohybů, snížení rozsahu a rychlosti pohybů, těžší cílená iniciace pohybů

Zadání úlohy	body
<p>Vyhledejte lineární vztahy mezi parametry.</p> <ul style="list-style-type: none"> Pomocí korelační analýzy najděte mezi možnými dvojicemi nezávislých parametrů takové, které budou vykazovat významný <u>lineární vztah</u>. Nezávislá dvojice například není celkové skóre dotazníku UPDRS a skóre z nějaké z jeho podsekcí. Hranici pro určení toho, který vztah je významný, a který ne, určete buď na základě vámi zvolené hraniční hodnoty korelačního koeficientu nebo pomocí statistické významnosti korelace vůči vybrané úrovni alfa. <p>Uveďte, jak jste nastavili hranici významnosti a korektně reportujte výsledky korelační analýzy pro nejvýznamnější dvojice.</p>	1
<ul style="list-style-type: none"> Vyberte jednu libovolnou významnou dvojici a vizualizujte jí ve 2D. Vytvořte lineární regresní model pro vybraná data. Parametry modelu (tj. v tomto případě směrnici a posun přímky) vypočítejte pomocí vztahu: $y = \mathbf{X} \cdot b \quad \text{kde: } y = [y_1, y_2, \dots, y_n]^T \quad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad b = [b_0, b_1]^T$ <ul style="list-style-type: none"> Váš výpočet srovnejte s výstupem funkce <code>fit</code> nebo <code>polyfit</code>. Pomocí vypočtených regresních parametrů si vykreslete lineární regresní přímku do obrázku s vizualizovanými daty. 	0.5
<p>Naprogramujte výpočet robustního lineárního model pomocí Theil-Senovy metody:</p> <ol style="list-style-type: none"> Pro všechny <u>unikátní</u> dvojice bodů z vybraného datasetu vypočítejte směrnici přímky k_i, která i-tou dvojicí prochází, tj.: $k_i = \frac{\Delta y}{\Delta x} = \frac{y_{1,i} - y_{2,i}}{x_{1,i} - x_{2,i}} \quad \text{přes všechny unikátní dvojice bodů: } i \in \langle 1, \frac{N(N-1)}{2} \rangle$ Vypočítejte medián \hat{k} ze všech směrnic. Výsledkem je směrnice výsledné robustní regresní přímky. Vypočítejte posun robustní regresní přímky \hat{b} (<i>bias, intercept</i>): $b_n = y_n - \hat{k} \cdot x_n \quad \hat{b} = \text{medián } b_n \text{ přes } n \in N$ <p>Pomocí vypočtených parametrů vykreslete robustní lineární regresní přímku do stejného obrázku jako v přechozím bodě. Odpovídá výsledek očekávání?</p>	1

Korelace UPDRS - UPDRS axial nebo UPDRS - UPDRS bradyk. nema cenu delat, protoze jsou to obviously zavisly data. Stejne tak asi bude AmpDec - VFT.
Nedelat Bonferroniho korekci pro velke mnozstvi testu.

Analyzujte, jak dobře vámi vytvořené lineární modely popisují výchozí data. Pro oba vaše modely (klasický lineární a robustní Theil-Senův) manuálně vypočtete následující „goodness-of-fit“ (GOF) veličiny:

1. **Dvouvzorkový Kolmogorov-Smirnov test** (`kstest2`): zapište hypotézy, které bude test ověřovat, výsledek korektně reportujte.
2. **SSE** (*Sum squared error*)
3. **RMSE** (*Root mean squared error*)
4. Koeficient determinace **R²**
5. **Adjusted R²**

$$RMSE = \sqrt{\left(\frac{SSE}{n-m}\right)} \quad R_{adj}^2 = 1 - \frac{(1-R^2)(n-1)}{n-m-1}$$

Ve vzorcích výše je n množství dat, na kterých je model založen (délka y) a m je počet parametrů, které jsou použity pro výpočet modelu (řád modelu).

Nakonec vytvořte pomocí funkce `fitlm` **vícedimenzionální lineární regresní model**, který bude mít stejnou výstupní proměnnou y a pro výpočet modelu použije všechny dostupné nezávislé parametry z tabulky. Nechte si vypsát výstup funkce `fitlm`, který obsahuje automaticky vypočtené GOF veličiny.

- Porovnejte výsledky GOF klasického a robustního modelu. Vysvětlete rozdíly mezi nimi.
- Vysvětlete rozdíly v GOF mezi modelem založeným na jednom parametru a vícedimenzionálním modelem. Jaké jsou výhody a nevýhody obou modelů?

1.5

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., Rusz, 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)

Krupička R., Krýže P., Neřuková S., Duspivová T., Klempíř O., Szabó Z., Dušek P., Šonka K., Rusz J. and Růžička E. (2020). *Instrumental analysis of finger tapping reveals a novel early biomarker of parkinsonism in idiopathic rapid eye movement sleep behaviour disorder*. Sleep Medicine, 75, 45-49, <https://doi.org/10.1016/j.sleep.2020.07.019>