**Příprava, zpracování a interpretace statistického zpracování dat**

**Normální rozdělení** pravděpodobnosti (normal probability distribution) je spojité **rozdělení** pravděpodobnosti, které popisuje celou řadu veličin, jejichž hodnoty se symetricky shlukují kolem střední hodnoty a vytvářejí tak charakteristický tvar hustoty pravděpodobnosti, která je známá také pod pojmem Gaussova křivka.

**Komparativní** – srovnávací. Srovnání dat na základě výsledků z minimálně dvou výzkumů. Je zde provedena komparace (porovnávání) získaných údajů a na tomto základě lze dospět

k novým poznatkům.

**Korelační –** asociace. V korelačních studiích se srovnává výskyt jevu ve vztahu k různým faktorům (věk, pohlaví, čas, konzumace určitých produktů, užívání léků apod. Mírou asociace mezi předpokládaným, např. rizikem a nemocí je [korelační koeficient,](https://www.wikiskripta.eu/w/M%C4%9B%C5%99en%C3%AD_z%C3%A1vislosti%2C_korelace_a_regrese) jehož hodnota se pohybuje od +1 do -1.

**ANOVA** - Analýza rozptylu (anglicky Analysis of variance) je metodou matematické statistiky, která umožňuje ověřit, zda na hodnotu náhodné veličiny pro určitého jedince má statisticky významný vliv hodnota některého znaku, který se u jedince dá pozorovat.

**O statistické významnosti** hovoříme tehdy, když nastane taková odchylka od teoretického očekávaní, která by za platnosti předem daného předpokladu měla velmi malou pravděpodobnost. V takovém případě se má za to, že předpoklad není správný. Statistická významnost se používá nejčastěji v souvislosti s testováním hypotéz.

**Velikost účinku** (anglicky **Effect size**) v matematické statistice je míra síly jevu, která je nezávislá na jednotkách měření. K běžně používaným statistikám tohoto typu patří:

* míry statistické závislosti spojitých proměnných jako korelační koeficient a koeficient determinace,
* míry závislosti nominálních znaků jako Pearsonův koeficient φ nebo Cramérovo V,
* míry rozdílu průměrů jako Cohenovo d a Hedgesovo g.

Předností velikosti účinku je to, že umožňuje snadno posoudit praktickou významnost statistické závislosti. Pokud však jednotky měření mají přirozený význam (např. počet denně odcvičených hodiny), doporučuje se vedle velikosti účinku ve zprávě uvádět i nestandardizované statistiky, například rozdíl v průměrném počtu odcvičených hodin denně vůči kontrolní skupině, a nikoli pouze velikost účinku.

**Načtení dat**

Načtení datového souboru Soubor -> Otevřít -> vybrat datový soubor -> Otevřít -> Importovat vybraný list do tabulky -> vybrat list Excelovského souboru -> OK -> nechat zatržené 1. řádek jako názvy proměnných -> OK -> Importovat jako textové popisky

Uložení datového souboru Soubor -> Uložit -> zadáme název souboru -> Uložit

Zapnutí automatického filtru Označit všechny sloupce (např. pomocí CTRL+A nebo kliknutím do levého horního rohu tabulky) -> Data -> Automatický filtr -> Automatický filtr

**Grafy vizualizace**

Vizualizace dat Vytváření grafů pomocí záložky Grafy. Koláčový graf Grafy -> 2D grafy -> Výsečové grafy -> zvolit proměnnou (např. Gender) (v záložce Detaily je možné zvolit, jakou legendu, typ a tvar grafu chceme (Legenda, Typ, Tvar) -> OK Po dvojím kliknutí na graf se nám ukáže okno Možnosti grafu, kde lze libovolně měnit barvu grafu i typ a tvar grafu a další parametry

Sloupcový graf (na ose y počty lidí) Grafy -> Histogramy -> Proměnné -> zvolit proměnnou (např. Group) -> OK -> zrušit zatržení Typ proložení: Normální -> na záložce Detaily zatrhnout Mezery mezi sloupci -> OK

Sloupcový graf (na ose y procenta) Grafy -> Histogramy -> Proměnné -> zvolit proměnnou (např. Group) -> OK -> zrušit zatržení Typ proložení: Normální -> na záložce Detaily zatrhnout Mezery mezi sloupci -> na záložce Detaily změnit u Osa Y hodnotu N na % -> OK

Histogram (na ose y procenta) Grafy -> Histogramy -> Proměnné -> zvolit proměnnou (např. Age) -> OK -> na záložce Detaily změnit u Osa Y hodnotu N na % (lze např. si vypsat i základní popisnou statistiku zatrhnutím Popisné statistiky) -> OK

Krabicový graf (s vykreslením odlehlých hodnot) Grafy -> 2D grafy -> Krabicové grafy... -> Proměnné -> zvolit proměnnou (např. Age) jako Závislé prom. -> OK -> OK

Krabicový graf (s minimem a maximem) Grafy -> 2D Grafy -> Krabicové grafy... -> Proměnné -> zvolit proměnnou (např. Age) jako Závislé prom. -> OK -> na záložce Detaily -> u Svorka zvolit Min-Max -> u Odl. hodn. & extrémy zvolit Vyp. -> OK

**Příprava dat pro analýzu**

Nastavení formátu u MMSE na double

Dvakrát kliknout na šedé políčko s názvem proměnné -> nastavit Typ na Double -> nastavit Formát zobrazení na Číslo -> OK

Nastavení formátu u scan\_date na datum

Dvakrát kliknout na šedé políčko s názvem proměnné -> nastavit Formát zobrazení na Datum -> vybrat formát 17/03/10 -> OK

Identifikace a odstranění duplikací

Data -> Filtrování dat/Překódování -> Filtrovat duplicitní případy -> Vstup: Proměnné -> ID -> OK -> u Output zatrhnout Vytvořit tabulku duplicit -> OK Je patrné, že se vždy zachová první záznam a druhý záznam je vyřazen bez ohledu na datum pořízení skenu. Pokud chceme, aby byl vždy odstraněn záznam se starším datem, je nejprve nutné data seřadit podle data pořízení skenu (sestupně) pomocí: Data -> Setřídit -> označit 1-ID -> Přidat prom. -> označit 30-scan\_date -> Přidat prom. -> Sestupně -> OK -> Zahrnout formátování Nový datový soubor bez duplikací uložit.

Odstranění chybějících a chybných hodnot

Data -> Podmnožina -> Případy -> zatrhnout Povolit podmínky výběru – zapnout filtr -> do Zadané výrazem napsat v4="" OR v4>110 OR v7="" -> OK -> OK Nový datový soubor bez chybějících a chybných hodnot uložit.

Rekódování proměnné Gender, aby obsahovala pouze hodnoty F a M

1. způsob – ručně: Vyfiltrovat si řádek s hodnotou FF a hodnotu FF přepsat na F

2. způsob – vytvořením nové proměnné: Označit proměnnou za proměnnou Gender -> Vložit -> Přidat proměnné -> Jméno -> zadat název nové proměnné (např. Gender\_rek) -> do kolonky Dlouhé jméno napsat =iif(v3="FF","F",v3) -> OK

Rekódování proměnné Group, aby obsahovala pouze hodnoty 1 (CN), 2 (MCI) a 3 (AD)

Označit proměnnou za proměnnou Group -> Vložit -> Přidat proměnné -> Jméno -> zadat název nové proměnné (např. Group\_3kat) -> do kolonky Dlouhé jméno napsat =iif(v2=3;2;iif(v2=4;3;v2)) -> OK Jiný způsob pomocí Data -> Překódovat...

Vytvoření textových popisků u kvalitativní proměnné

Dvakrát kliknout na šedé políčko s názvem proměnné -> Textové hodnoty -> zadat textové popisky a jejich příslušné číselné hodnoty -> OK -> OK

**Popisná sumarizace dat**

Popisná sumarizace dat pomocí Statistiky -> Základní statistiky/tabulky. Obecný popis dialogového okna pro sumarizaci dat, vizualizace a další analýzy je uveden na Obr. 1.

Popisná sumarizace kvalitativních dat – frekvenční tabulka Statistiky -> Základní statistiky/tabulky -> Tabulky četností -> Proměnné -> zvolit proměnnou (např. Group) -> OK -> Výpočet

Popisná sumarizace kvantitativních dat Statistiky -> Základní statistiky/tabulky -> Popisné statistiky -> Proměnné -> zvolit proměnnou (např. Age) -> OK -> na záložce Detailní výsledky zatrhnout Medián, Variační koeficient, Dolní & horní kvartily -> Výpočet

**Intervaly spolehlivosti**

Výpočet intervalu spolehlivosti a střední chyby průměru (standard error) Statistiky -> Základní statistiky/tabulky -> Popisné statistiky -> Proměnné -> zvolit proměnnou (např. Age) -> OK -> na záložce Detailní výsledky zatrhnout Meze spolehl. prům. a Sm. chyba průměru -> Výpočet

Výpočet kvantilů Studentova rozložení Statistiky -> Pravděpodobnostní kalkulátor -> Rozdělení... -> t (Studentovo) -> zatrhnout Inverze -> jako p zadat 0.975 -> jako sv (počet stupňů volnosti) zadat 832 -> Výpočet (vypočítá nám to hodnotu t)

**Výpočet velikosti vzorku a power analýza**

Výpočet velikosti vzorku Statistiky -> Analýza síly testu -> Výpočet velikosti vzorku -> zvolit typ výpočtu podle typu našeho experimentu, který budeme chtít provést (např. Dva průměry, t-test, nezáv. vzorky) -> OK -> nastavit parametry a zvolit typ hypotézy (např. viz Obr. 3 – Pozor! Zadáváme typ nulové hypotézy, tedy zadáváme opak toho, co chceme prokázat!) -> OK -> Vypočítat N

**Power analýza**

Statistiky -> Analýza síly testu -> Výpočet síly testu -> zvolit typ výpočtu podle typu našeho experimentu, který budeme chtít provést (např. Dva průměry, t-test, nezáv. vzorky) -> OK -> nastavit parametry a zvolit typ hypotézy (např. viz Obr. 4 – Pozor! Zadáváme typ nulové hypotézy, tedy zadáváme opak toho, co chceme prokázat!) -> OK -> Vypočítat sílu

**Statistické testy pro kvantitativní data – parametrické testy**

Jednovýběrový t-test

Statistiky -> Základní statistiky/tabulky -> t-test, samost. vzorek -> OK -> zvolit proměnnou (např. Hippocampus\_volume (mm3)) -> OK -> Test všech průměrů vůči: 6575 -> na záložce Možnosti zatrhnout Výpočet mezí spolehl. -> Výpočet

Párový t-test

Statistiky -> Základní statistiky/tabulky -> t-test, závislé vzorky -> OK -> zvolit proměnné (např. Hippocampus\_volume (mm3) jako 1. seznam proměnných a Hippocampus\_volume\_24 (mm3) jako 2. seznam proměnných) -> OK -> Výpočet

Dvouvýběrový t-test

Statistiky -> Základní statistiky/tabulky -> t-test, nezávislé, dle skupin -> OK -> zvolit proměnné (např. Putamen\_volume (mm3) jako Závislé proměnné a Gender\_rek jako Grupovací proměnná) -> na záložce Možnosti lze zvolit Levenův test (test homogenity rozptylů) a Meze spol. pro odhady -> Výpočet

Analýza rozptylu (ANOVA) a post-hoc testy –

1. způsob Statistiky -> Základní statistiky/tabulky -> Rozklad & jednofakt. ANOVA -> OK -> zvolit proměnné (např. Hippocampus\_volume (mm3) jako Závislé proměnné a Group\_3kat jako Grupovací proměnné) -> OK -> OK -> na záložce ANOVA & testy kliknout na Analýza rozptylu (vypíše ANOVA tabulku); dále lze vypsat i výsledky testů homogenity rozptylů: Leveneovy testy, Brown-Forsytheho testy -> na záložce Post-hoc kliknout na Tukeyův HSD (v případě vyrovnaných počtů subjektů ve skupinách), Tukey HSD - nestejná N (v případě nestejných počtů subjektů ve skupinách) nebo Schefféův test (pro stejné i nestejné počty subjektů ve skupinách) Analýza rozptylu (ANOVA) a post-hoc testy –

2. způsob Statistiky -> ANOVA -> Jednofaktorová ANOVA -> OK -> zvolit proměnné (např. Hippocampus\_volume (mm3) jako Seznam závislých proměnných a Group\_3kat jako Kategor. nezávislá proměnná (faktor)) -> OK -> OK -> Všechny efekty (vypíše ANOVA tabulku; prvního řádku s interceptem si nevšímáme) -> Více výsledků -> na záložce Post-hoc kliknout na Tukeyův HSD (pro stejné počty subjektů ve skupinách), HSD nestejné N (pro nestejné počty subjektů ve skupinách) nebo Schefféův (pro stejné i nestejné počty subjektů)

**Statistické testy pro kvantitativní data – neparametrické testy**

Wilcoxonův test – jednovýběrový

STATISTICA neumožňuje počítat jednovýběrový Wilcoxonův test přímo. Je nutné nejprve vytvořit novou proměnnou, která bude mít ve všech řádcích hodnotu, se kterou chceme srovnávat naše data: Vložit -> Přidat proměnné -> Jméno -> zadat název nové proměnné (např. mmse\_konst) -> do kolonky Dlouhé jméno napsat =27,5 (hodnota konstanty, se kterou chceme srovnávat) -> OK Poté můžeme použít pro výpočet párový Wilcoxonův test: Statistiky -> Neparametrická statistika -> Porovnání dvou závislých vzorků (proměnné) -> OK -> zvolit proměnné (např. MMSE jako 1. seznam proměnných a mmse\_konst jako 2. seznam proměnných) -> OK -> Wilcoxonův párový test (Je možné vypočítat i Znaménkový test, který je též neparametrickou alternativou párového t-testu.)

Wilcoxonův test – párový

Statistiky -> Neparametrická statistika -> Porovnání dvou závislých vzorků (proměnné) -> OK -> zvolit proměnné (např. MMSE jako 1. seznam proměnných a MMSE\_24 jako 2. seznam proměnných) -> OK -> Wilcoxonův párový test (Je možné vypočítat i Znaménkový test, který je též neparametrickou alternativou párového t-testu.)

Mannův-Whitneyův test

Statistiky -> Neparametrická statistika -> Porovnání dvou nezávislých vzorků (skupiny) -> OK -> zvolit proměnné (např. Hippocampus\_volume (mm3) jako Seznam závislých proměnných a Gender\_rek jako Nezáv. (groupov.) proměnná) -> OK -> M-W U test

Kruskalův-Wallisův test

Statistiky -> Neparametrická statistika -> Porovnání více nezávislých vzorků (skupiny) -> OK -> zvolit proměnné (např. MMSE jako Seznam závislých proměnných a Group\_3kat jako Nezáv. (groupov.) proměnná) -> OK -> Výpočet (vypíše výsledky Kruskalova-Wallisova testu) -> Vícenás. porovnání průměrného pořadí pro vš. sk. (vypíše výsledky post hoc analýzy)

**Ověření předpokladů statistických testů**

Vykreslení Q-Q grafu pro jednotlivé skupiny

Grafy -> 2D grafy -> Normální pravděpodobnostní grafy... -> zvolit proměnnou (např. Hippocampus\_volume (mm3)) -> Anal. skup. -> Skup. proměnná(é) -> vybrat proměnnou (např. Gender\_rek) -> OK -> zatrhnout Výstup do jediné složky -> přepnout Uspořádání skupin na Sestupně -> OK -> lze zatrhnout Shapiro-Wilksův test (test normality dat) -> OK

Ověření normality pomocí Základní statistiky/tabulky pro jednotlivé skupiny

Statistiky -> Základní statistiky/tabulky -> Popisné statistiky -> Proměnné -> zvolit proměnnou (např. Hippocampus\_volume (mm3)) -> OK -> na záložce Detailní výsledky zatrhnout Medián (to teď není nutné) -> Anal. skup. -> Skup. proměnná(é) -> vybrat proměnnou (např. Gender\_rek) -> OK -> zatrhnout Výstup do jediné složky a Sloučit tabulkové výsledky v jedné tabulce -> OK -> přepnout Uspořádání skupin na Sestupně -> OK -> na záložce Normalita zatrhnout Shapiro-Wilkův W test -> Tabulky četností -> na záložce Pravd. & bod. grafy lze nechat vykreslit Normální pravděpod. graf (tzn. Q-Q graf) -> na záložce Normalita lze nechat vykreslit Histogramy

Ověření normality pomocí Histogramy... pro jednotlivé skupiny

Grafy -> Histogramy -> Proměnné -> zvolit proměnnou (např. Hippocampus\_volume (mm3)) -> OK -> na záložce Detaily změnit u Osa Y hodnotu N na %, zatrhnout Shapiro-Wilksův test a KolmogorovSmirnovův test -> Anal. skup. -> Skup. proměnná(é) -> vybrat proměnnou (např. Gender\_rek) -> OK -> zatrhnout Výstup do jediné složky a Sloučit tabulkové výsledky v jedné tabulce -> OK -> přepnout Uspořádání skupin na Sestupně -> OK -> OK

**Analýza kontingenčních tabulek**

Kontingenční tabulka absolutních četností

Statistiky -> Základní statistiky/tabulky -> Kontingenční tabulky -> OK -> Specif. tabulky (vyberte proměn.) (např. Group\_3kat do List 1 a Age\_kat do List 2)-> OK -> OK -> Výpočet

Kontingenční tabulka procent

 Statistiky -> Základní statistiky/tabulky -> Kontingenční tabulky -> OK -> Specif. tabulky (vyberte proměn.) (např. Group\_3kat do List 1 a Age\_kat do List 2)-> OK -> OK -> na záložce Možnosti zatrhnout Procenta z počtu v řádku, Procenta z počtu ve sloupci nebo Procenta celkového počtu -> Výpočet

Kontingenční tabulka – očekávané četnosti

Statistiky -> Základní statistiky/tabulky -> Kontingenční tabulky -> OK -> Specif. tabulky (vyberte proměn.) (např. Group\_3kat do List 1 a Age\_kat do List 2)-> OK -> OK -> na záložce Možnosti zatrhnout Očekávané četnosti -> Výpočet

Pearsonův chí-kvadrát test

Statistiky -> Základní statistiky/tabulky -> Kontingenční tabulky -> OK -> Specif. tabulky (vyberte proměn.) (např. Group\_3kat do List 1 a Age\_kat do List 2)-> OK -> OK -> na záložce Možnosti zatrhnout Pearsonův & M-V chí-kvadrát -> na záložce Detailní výsledky kliknout na Detailní 2-rozm. tabulky

Fisherův exaktní test

Statistiky -> Základní statistiky/tabulky -> Kontingenční tabulky -> OK -> Specif. tabulky (vyberte proměn.) (např. Gender do List 1 a mmse\_kat do List 2)-> OK -> OK -> na záložce Možnosti zatrhnout Fisher exakt., Yates, McNemar (2 x 2) -> na záložce Detailní výsledky kliknout na Detailní 2-rozm. tabulky

McNemarův test

Statistiky -> Základní statistiky/tabulky -> Kontingenční tabulky -> OK -> Specif. tabulky (vyberte proměn.) (např. mmse\_kat do List 1 a mmse24\_kat do List 2) -> OK -> OK -> na záložce Možnosti zatrhnout Fisher exact, Yates, McNemar (2 x 2) -> na záložce Detailní výsledky kliknout na Detailní 2-rozm. Tabulky

**Korelační analýza**

Bodový graf (“Scatterplot”)

Grafy -> Bodové grafy... -> zvolit proměnné (např. Putamen\_volume (mm3) jako X a Amygdala\_volume (mm3) jako Y -> OK -> na záložce Detaily lze zatrhnout Korelace a p (lin. prolož.) (vypočítá Pearsonův korelační koeficient a p-hodnotu) a R kvadrát -> OK

Výpočet Pearsonova korelačního koeficientu

Statistiky -> Základní statistiky/tabulky -> Korelační matice -> OK -> 1 seznam proměn. -> zvolit proměnné (např. Amygdala\_volume (mm3) a Putamen\_volume (mm3)) -> na záložce Možnosti zvolit Zobrazit r, p-hodnoty a N -> Výpočet

Srovnání dvou korelačních koeficientů

Statistiky -> Základní statistiky/tabulky -> Testy rozdílů: r, %, průměry -> OK -> zadáme hodnoty korelačního koeficientu a počet subjektů obou výběrů -> Výpočet (spočítá nám to p-hodnotu)

Výpočet Spearmanova korelačního koeficientu (výpočet čtvercové korelační matice)

Statistiky -> Neparametrická statistika -> Korelace (Spearman, Kendallovo tau, gama) -> OK -> zvolit proměnné (např. MMSE a Hippocampus\_volume (mm3)) -> OK -> Spearmanův koef. R

Výpočet Spearmanova korelačního koeficientu (výpočet detailní tabulky)

Statistiky -> Neparametrická statistika -> Korelace (Spearman, Kendallovo tau, gama) -> OK -> zvolit Detailní report (místo Čtvercová matice) -> zvolit proměnné (např. MMSE do 1. seznam proměnných a Hippocampus\_volume (mm3) do 2. seznam proměnných) -> OK -> Spearmanův koef. R

**Regresní analýza**

Lineární regrese a odstranění vlivu kovariát Statistiky -> Vícenásobná regrese -> zvolit proměnné (např. Nucl\_caud\_volume (mm3) jako Závislá prom., Age a gender\_01,... jako Seznam nezáv. proměnných) -> OK -> OK -> Výpočet: Výsledky regrese (vypíše regresní koeficienty a p-hodnoty) -> OK

- Vykreslení Q-Q grafu pro rezidua: kliknout na Normální p-graf reziduí

- Vykreslení histogramu reziduí: na záložce Rezidua kliknout na Histogram reziduí (vpravo lze zvolit, zda chceme vykreslit histogram reziduí či standardizovaných reziduí)

- Vykreslení bodového grafu predikovaných hodnot a reziduí: na záložce Bodové grafy kliknout na Předpovědi vs. rezidua

- Uložení reziduí: na záložce Uložit kliknout na Uložit rezidua & předpovědi -> zvolit proměnné, které bude nově vytvořená tabulka dále obsahovat -> OK