

Analýza dat pro Neurovědy



RNDr. Eva Janoušová
doc. RNDr. Ladislav Dušek, Dr.

Blok 4

Jak a kdy použít parametrické a
neparametrické testy II.

Osnova

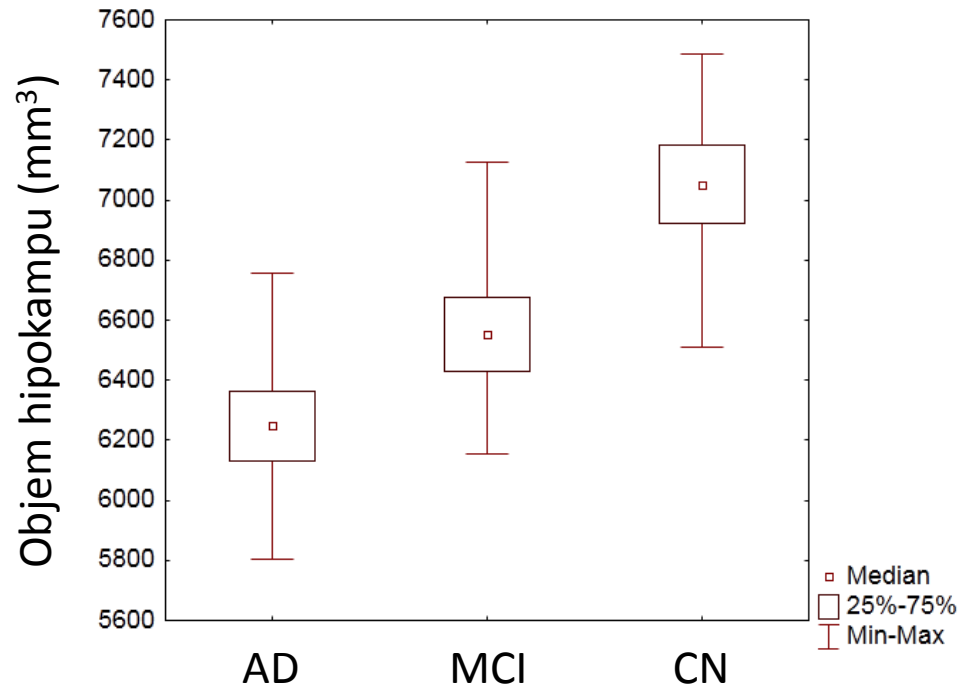
1. Analýza rozptylu (ANOVA)
2. Problém násobného testování hypotéz a použití korekčních procedur
3. Kruskalův-Wallisův test
4. Analýza rozptylu jako lineární model

Parametrické a neparametrické testy pro kvantitativní data – přehled

Typ srovnání	Parametrický test	Neparametrický test
1 skupina dat s referenční hodnotou – jednovýběrové testy:	Jednovýběrový t-test, jednovýběrový z-test	Wilcoxonův test
2 skupiny dat párově – párové testy:	Párový t-test	Wilcoxonův test, znaménkový test
2 skupiny dat nepárově – dvouvýběrové testy:	Dvouvýběrový t-test	Mannův-Whitneyův test, mediánový test
Více skupin nepárově:	ANOVA	Kruskalův- Wallisův test

1. Analýza rozptylu (ANOVA)

Motivace



Jak můžeme ověřit, zda se liší objem hipokampu u pacientů s AD, pacientů s MCI a u zdravých kontrol?

- Můžeme použít vhodný test pro dva výběry (např. dvouvýběrový t-test) a otestovat, jak se liší AD od MCI, AD od CN a MCI od CN – tedy provést 3 testy.
- Můžeme použít vhodný test pro více než dvě srovnávané skupiny.

V čem je zásadní rozdíl mezi A a B?

Motivace – pokračování

- **Problém s možností A je v násobném testování hypotéz:**

S narůstajícím počtem testovaných hypotéz nám roste také pravděpodobnost získání falešně pozitivního výsledku, tedy pravděpodobnost toho, že se při našem testování zmýlíme a ukážeme na statisticky významný rozdíl tam, kde ve skutečnosti žádný neexistuje (chyba I. druhu).

- Máme tři testy, v každém 95% pravděpodobnost, že neuděláme chybu I. druhu.
- Pro všechny tři testy to tedy znamená: $0,95 \times 0,95 \times 0,95 = 0,857$.
- Pravděpodobnost, že neuděláme chybu I. druhu nám celkově klesla na 0,857.
- **Pravděpodobnost, že uděláme chybu I. druhu nám celkově stoupla na 0,143.**

Motivace – pokračování

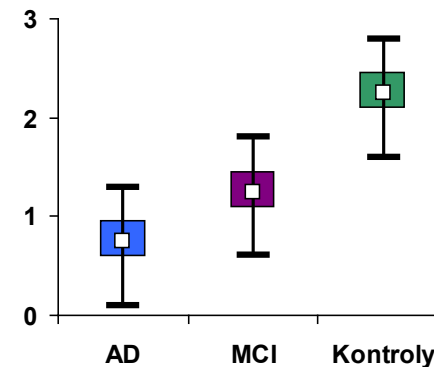
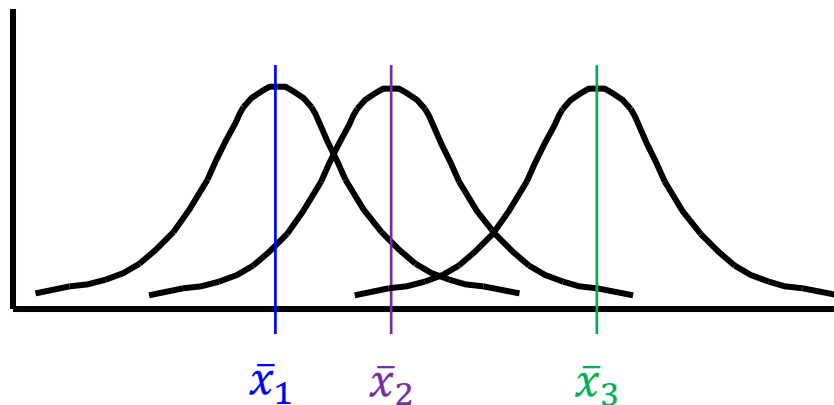
- Lepší volbou je:

B. Použít vhodný test pro více než dvě srovnávané skupiny.

- **Analýza rozptylu (ANOVA = „ANalysis Of VAriance“) je statistickou metodou, která umožňuje testovat rozdíl v průměrech více než dvou skupin. Přitom se jedná o jeden test.**

Analýza rozptylu (ANOVA) jednoduchého třídění

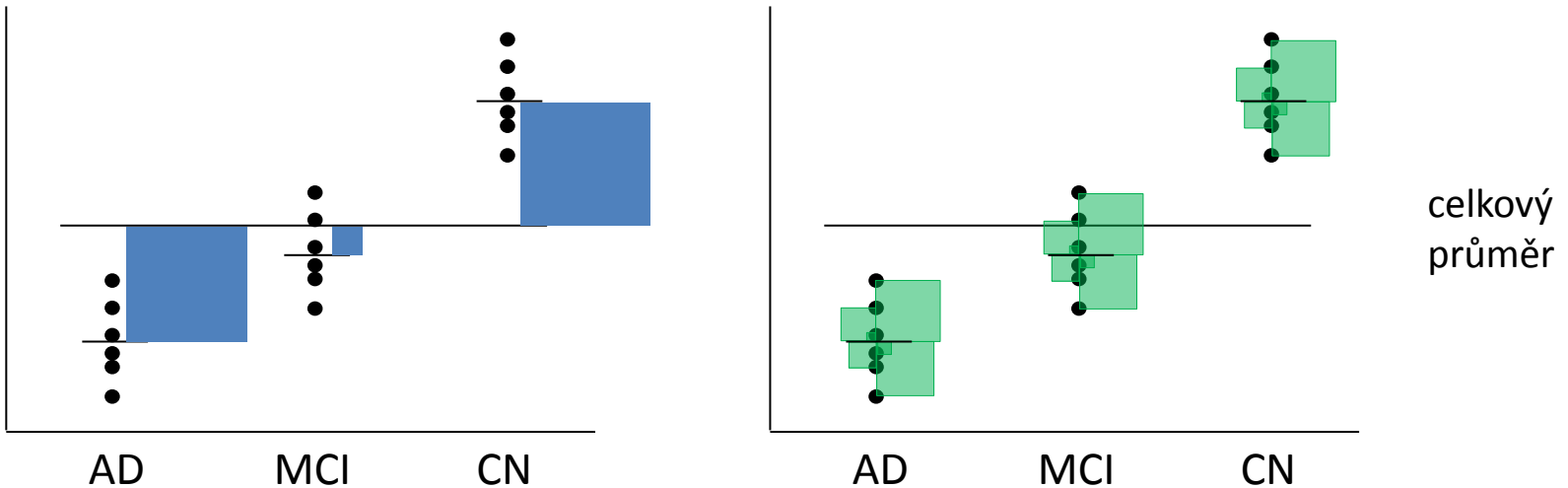
- Srovnáváme tři a více skupin dat, které jsou na sobě nezávislé (mezi objekty neexistuje vazba).
- Příklady: srovnání objemu hipokampu u pacientů s AD, pacientů s MCI a kontrol; srovnání kognitivního výkonu podle čtyř kategorií věku.



- Předpoklady: **normalita dat ve VŠECH skupinách, shodnost (homogenita) rozptylů VŠECH srovnávaných skupin**, nezávislost jednotlivých pozorování.
- Testová statistika: $F = \frac{S_A / df_A}{S_e / df_e}$ - vysvětlení na dalších slidech

Analýza rozptylu (ANOVA) – princip

- Srovnání variability (rozptylu) mezi výběry s variabilitou uvnitř výběrů.

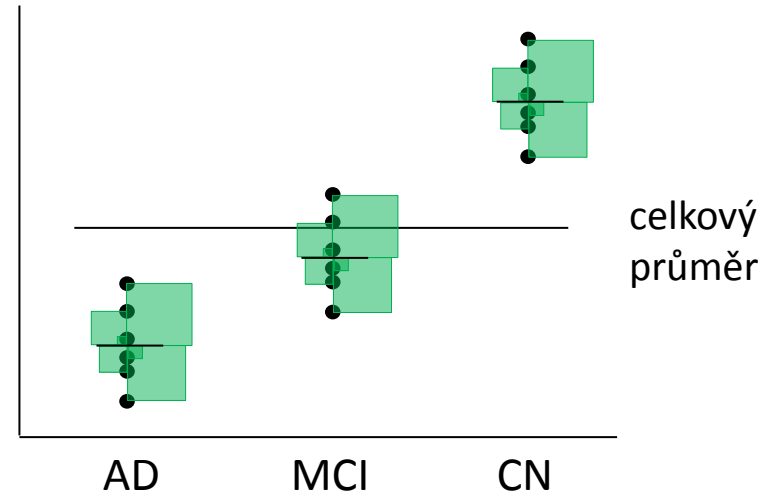
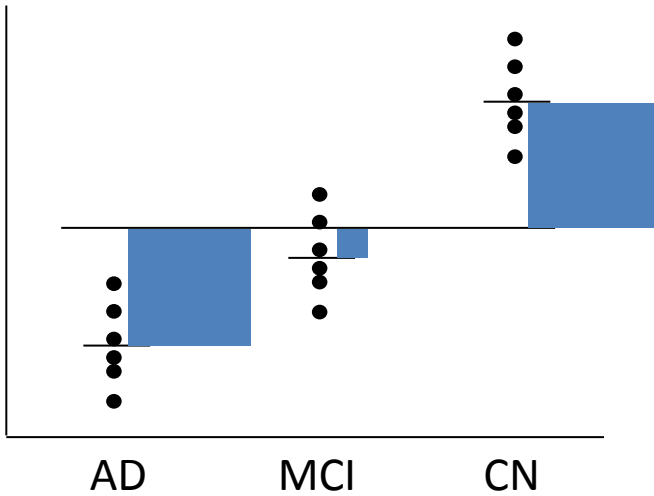


- Tabulka analýzy rozptylu jednoduchého třídění (One-Way ANOVA):

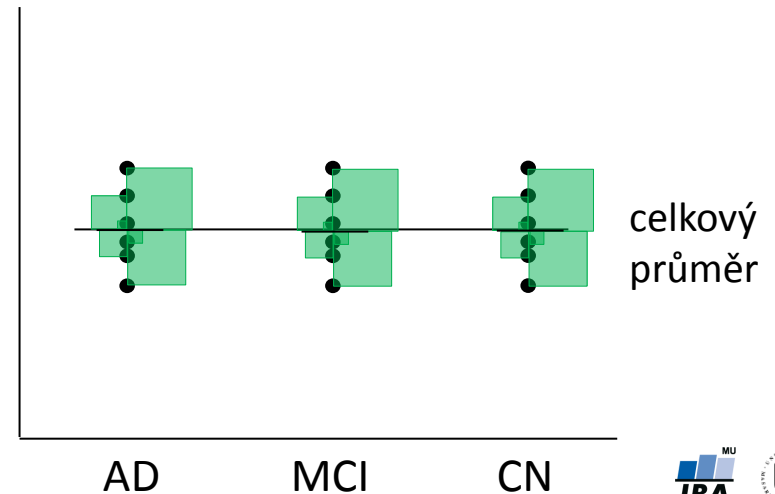
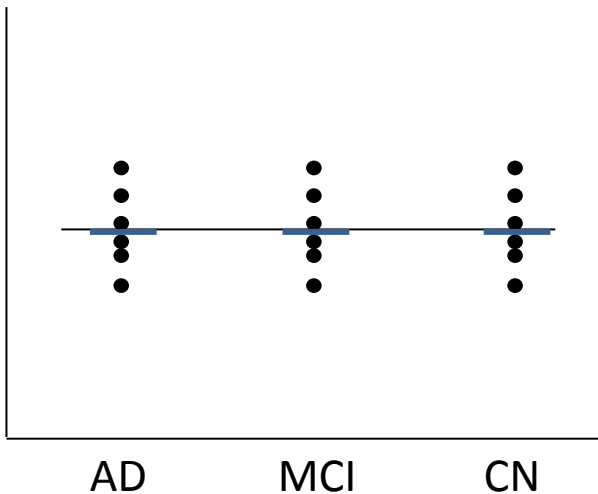
Variabilita	Součet čtverců	Počet stupňů volnosti	Průměrný čtverec	F statistika	p-hodnota
Mezi skupinami	S_A	$df_A = k - 1$	$MS_A = S_A / df_A$	$F = \frac{S_A / df_A}{S_e / df_e}$	p
Uvnitř skupin (reziduální var.)	S_e	$df_e = n - k$	$MS_e = S_e / df_e$		
Celkem	S_T	$df_T = n - 1$			

ANOVA – 2 ukázkové situace

- Rozdíl ve všech třech skupinách:



- Žádný rozdíl mezi skupinami:



Analýza rozptylu (ANOVA) jednoduchého třídění

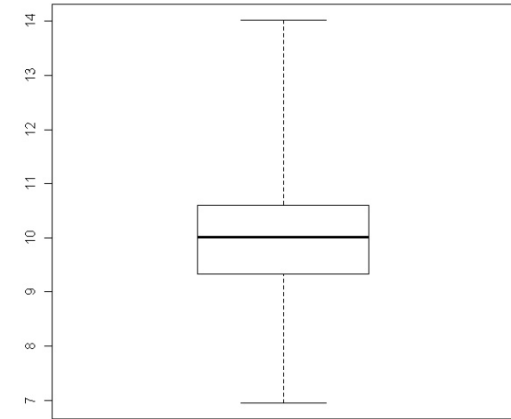
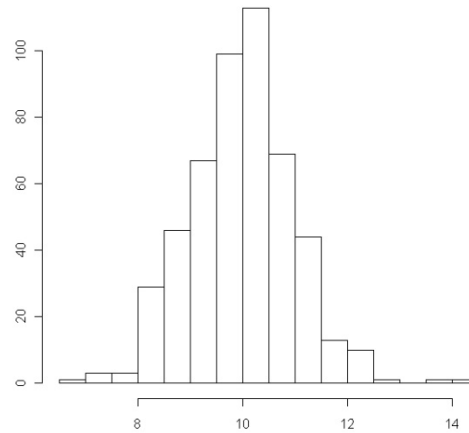
- **Příklad:** Chceme srovnat, zda se liší objem hipokampu podle typu onemocnění (tzn. u pacientů s AD, pacientů s MCI a zdravých kontrol).
- Tzn. hypotézy budou mít tvar: $H_0 : \mu_{AD} = \mu_{MCI} = \mu_{CN}$
 $H_1 : \text{nejméně jedno } \mu_i \text{ je odlišné od ostatních}$
- **Postup:**
 1. Popisná sumarizace objemu hipokampu podle typu onemocnění.
 2. Ověření normality hodnot ve VŠECH skupinách.
 3. Ověření shodnosti rozptylů VŠECH skupin.
 4. Aplikujeme statistický test.
 5. Nulovou hypotézu zamítneme nebo nezamítneme:
 $p < 0,001 < 0,05$ → zamítáme nulovou hypotézu → Rozdíl v objemu hipokampu podle typu onemocnění je statisticky významný (na hladině významnosti $\alpha=0,05$.)

Ověření normality dat

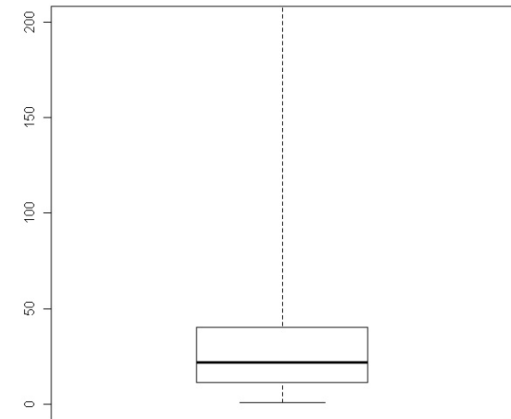
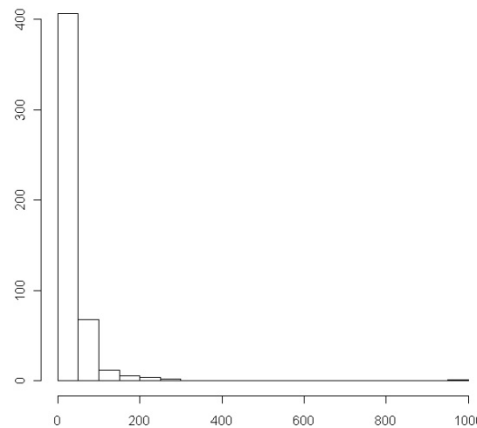
- **Graficky:**
 - histogram
 - krabicový graf (box-plot)
 - Q-Q graf
- **Testy normality:**
 - Shapirův-Wilkův test
 - Kolmogorovův-Smirnovův test
- **Testy nejsou vždy nejlepším nástrojem! Vždy je důležité se podívat i očima!**
- Pokud o sledované veličině prokazatelně víme, že v cílové populaci nabývá normální rozdělení (např. výška lidské postavy), ale v daném souboru normální rozdělení nepotvrdíme, **pak s naším náhodným výběrem není něco v pořádku** – např. není reprezentativní.

Ověření normality graficky – krabicový graf a histogram

- Normální rozdělení

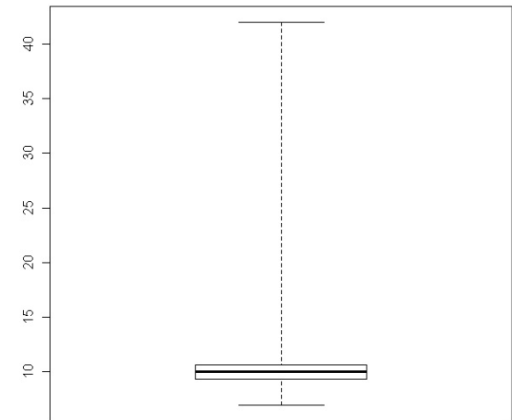
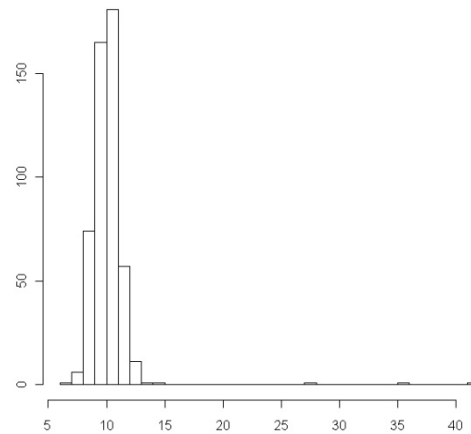


- Log-normální rozdělení

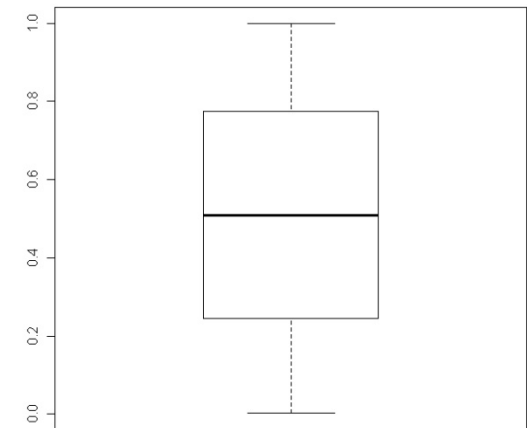
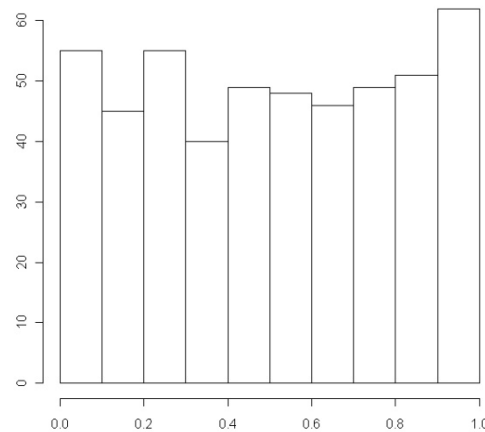


Ověření normality graficky – krabicový graf a histogram

- Normální rozdělení s odlehlými hodnotami

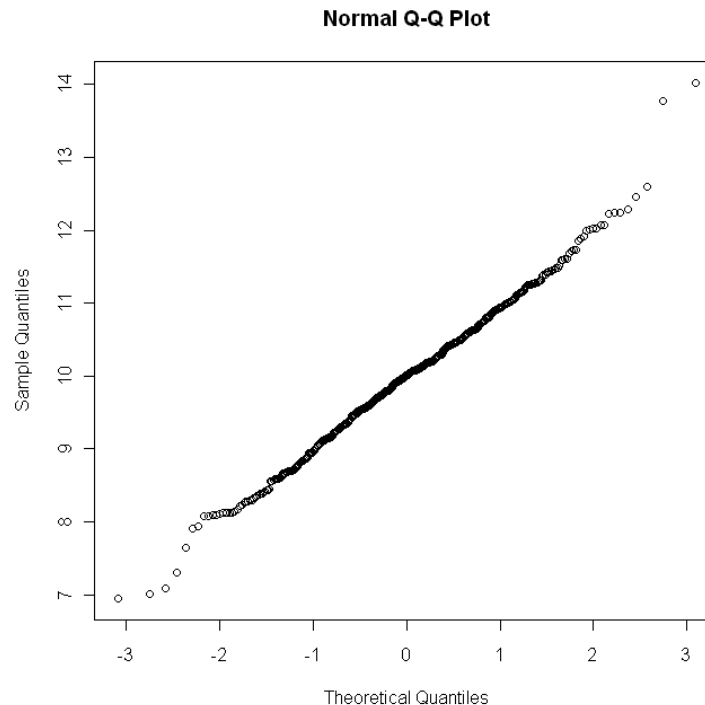


- Rovnoměrně spojité rozdělení



Ověření normality graficky – Q-Q graf

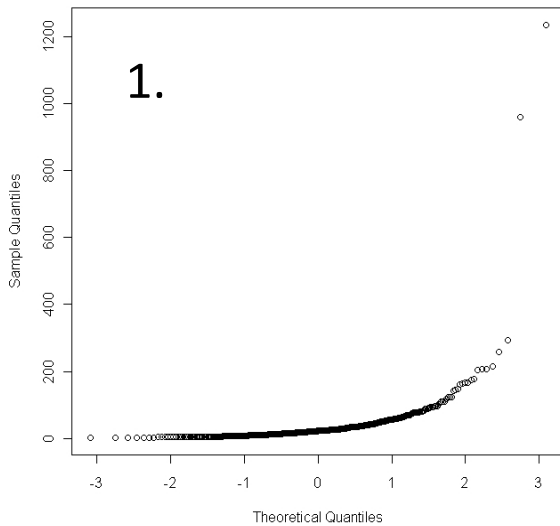
- Q-Q graf proti sobě zobrazuje kvantily pozorovaných hodnot a kvantily teoretického rozdělení pravděpodobnosti (zde normálního rozdělení).
- V případě shody leží všechny body na přímce.
- Normální rozdělení:



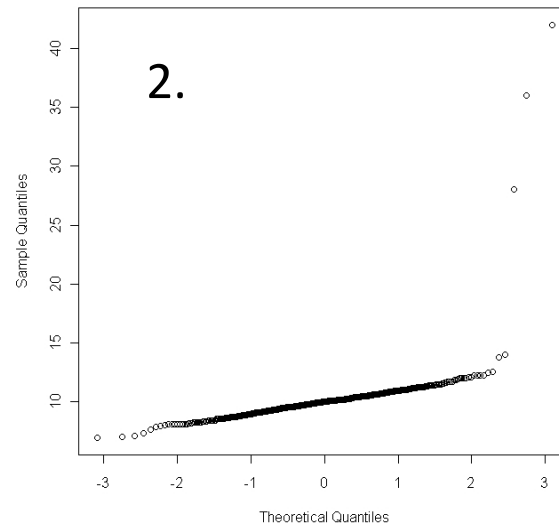
Ověření normality graficky – Q-Q graf

1. Log-normální rozdělení
2. Normální rozdělení s odlehlými hodnotami
3. Rovnoměrně spojité rozdělení

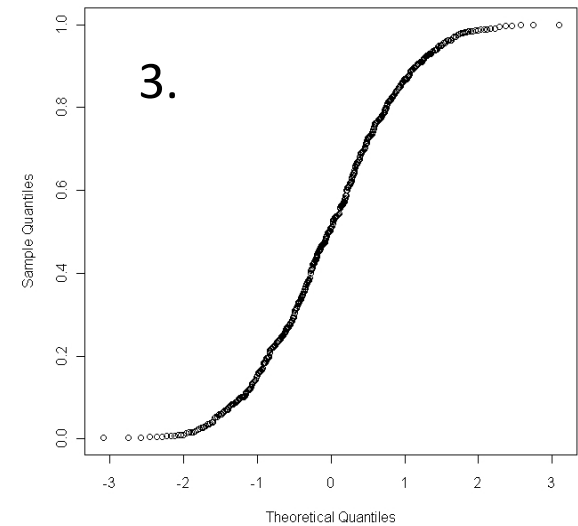
Normal Q-Q Plot



Normal Q-Q Plot



Normal Q-Q Plot

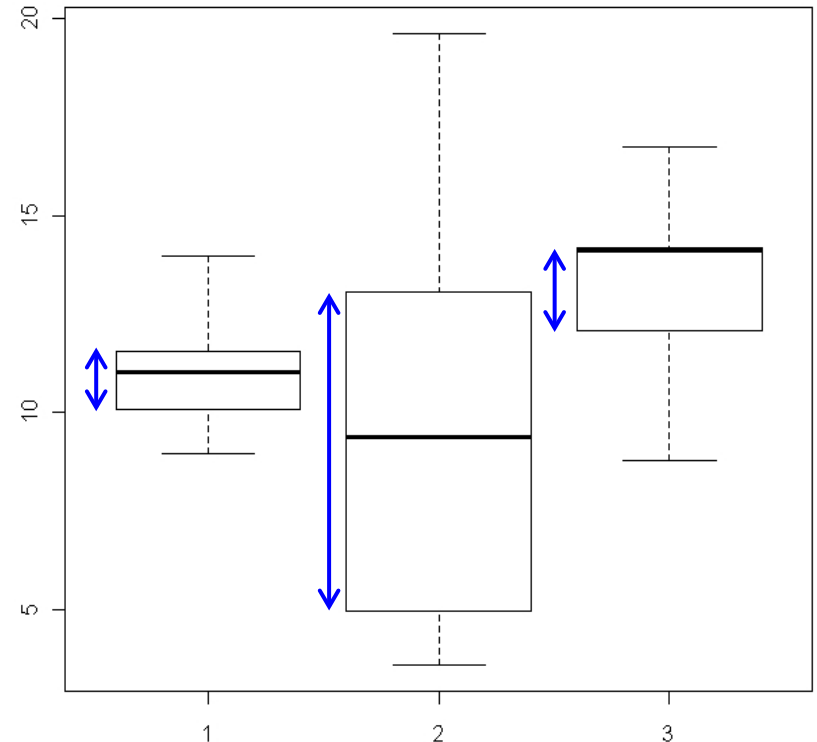


Ověření normality pomocí testů

- **Shapirův-Wilkův test** – v podstatě se jedná o proložení seřazených hodnot regresní přímkou vzhledem k očekávaným hodnotám normálního rozdělení. Má tedy přímý vztah k Q-Q plotu – vyhodnocuje, jak moc se Q-Q plot liší od ideální přímky. **Doporučován pro menší vzorky, může být „moc“ přísný pro velké vzorky.**
- **Kolmogorovův-Smirnovův test** – založen na srovnání výběrové distribuční funkce s teoretickou distribuční funkcí odpovídající normálnímu rozdělení. K-S test hodnotí maximální vzdálenost mezi těmito dvěma distribučními funkcemi. V praxi se používá korekce dle Lillieforse.

Ověření shody (homogeneity) rozptylů

- **Grafické ověření** – krabicový graf, histogram.
- **F-test** (testování shody rozptylů dvou vzorků)
- **Leveneův test** – často používaný (testování shody rozptylů dvou a více vzorků)
- **Bartlettův test**



Výsledky ANOVA testu

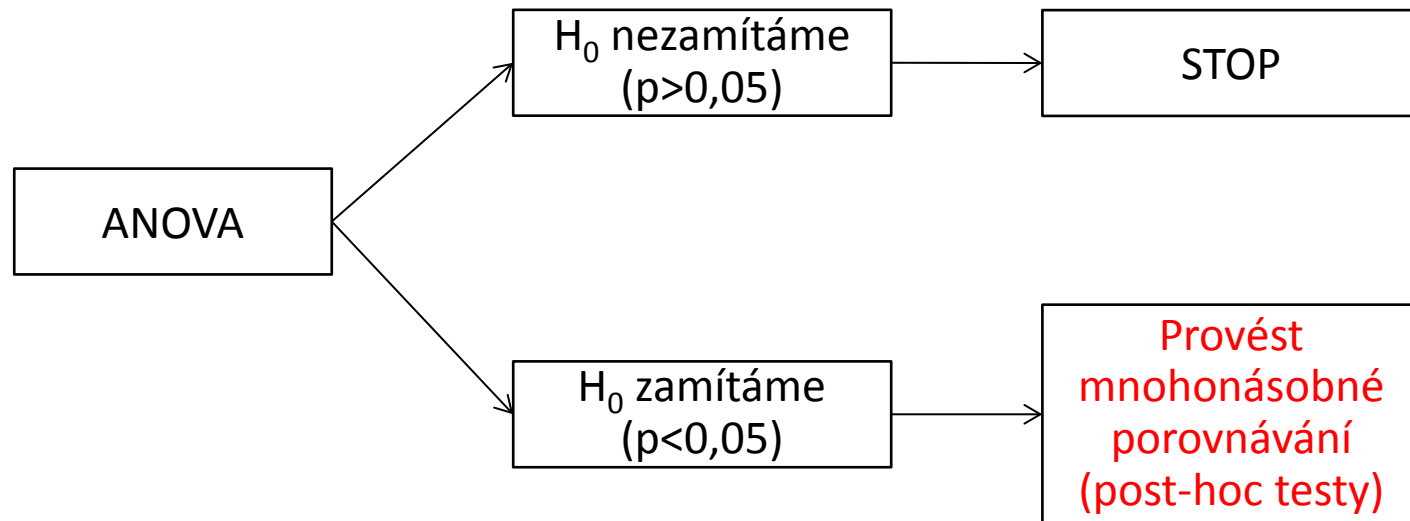
- Tabulka analýzy rozptylu jednoduchého třídění:

Variabilita	Součet čtverců	Počet stupňů volnosti	Průměrný čtverec	F statistika	p-hodnota
Mezi skupinami	$S_A =$ 71 422 222	$df_A = k - 1 =$ 2	$MS_A = S_A / df_A =$ 35 711 111	$F = \frac{S_A / df_A}{S_e / df_e} = 1103,6$	0,00
Uvnitř skupin (reziduální var.)	$S_e =$ 26 857 142	$df_e = n - k =$ 830	$MS_e = S_e / df_e =$ 32 358		
Celkem	$S_T =$ 98 279 364	$df_T = n - 1 =$ 832			

- Výsledek ze softwaru STATISTICA:

Analysis of Variance (Data_neuro_vycistena2)								
Marked effects are significant at p < ,05000								
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
Hippocampus volume (mm3)	71422222	2	35711111	26857142	830	32358,00	1103,625	0,00

Další kroky analýzy



2. Problém násobného testování hypotéz a použití korekčních procedur

Korekce na násobné srovnání výběrů

- Zamítneme-li analýzou rozptylu nulovou hypotézu o celkové rovnosti středních hodnot, má smysl se ptát, jaké skupiny se od sebe nejvíce liší.
- Toto srovnání lze provést pomocí testů pro dva výběry, ale je nutné korigovat výslednou hladinu významnosti testu, abychom se vyhnuli chybě I. druhu.
- Nejjednodušší metoda: **Boferroniho korekce** - korekce hladiny významnosti: $\alpha^* = \alpha/m$, kde m je počet provedených testů. Ekvivalentně lze vynásobit p -hodnotu počtem provedených testů. Nevýhodou je, že je konzervativní („přísná“) pro velké m , tedy počet provedených testů.
- Pro analýzu rozptylu: **Tukeyho** a **Scheffého post hoc testy**.
- Může se stát, že při použití různých korekcí nám mohou vyjít výsledky různě (např. při použití Scheffého testu nám vyjde statisticky významný rozdíl mezi skupinou AD a MCI a při použití Tukeyho testu nám rozdíl statisticky významný nevyjde).

Poznámka

- Může nastat situace, kdy zamítneme H_0 u ANOVY, ale metodami mnohonásobného porovnávání nenajdeme významný rozdíl u žádné dvojice středních hodnot. K tomu dochází zvláště tehdy, když p-hodnota pro ANOVU je jen o málo nižší než zvolená hladina významnosti.
- Důvod: post-hoc testy (tzn. metody mnohonásobného porovnávání) mají obecně menší sílu než ANOVA, proto nemusí odhalit žádný rozdíl.

Korekce na násobné srovnání – jiná situace

- Problém násobného testování („Multiple Testing Problem“) nastává, i když je provedeno **větší množství testů na různých proměnných** v rámci jednoho hodnocení dat.
- Příklad: zjišťování, zda se liší objem šedé hmoty u dvou skupin subjektů v každém voxelu obrazu.
- Korekce:
 - **Bonferroniho korekce** – kontroluje pravděpodobnost, s jakou dostaneme falešně pozitivní výsledek (kontroluje chybu I. druhu); konzervativní pro velký počet provedených testů.
 - **False discovery rate (FDR)** – kontroluje podíl falešně pozitivních výsledků mezi všemi statisticky významnými výsledky (např. pokud je FDR 0,05 a počet všech statisticky významných výsledků bude 1000, tak můžeme očekávat, že 50 výsledků bude falešně pozitivních).

Úkol 1.

- **Zadání:** Zjistěte, zda se liší objem pallida podle typu onemocnění (nezapomeňte ověřit předpoklady).
- **Řešení:**

Analysis of Variance (Data_neuro_vycistena2)								
Marked effects are significant at $p < ,05000$								
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
Pallidum_volume (mm3)	229575,6	2	114787,8	34702692	830	41810,47	2,745432	0,064804

Parametrické a neparametrické testy pro kvantitativní data – přehled

Typ srovnání	Parametrický test	Neparametrický test
1 skupina dat s referenční hodnotou – jednovýběrové testy:	Jednovýběrový t-test, jednovýběrový z-test	Wilcoxonův test
2 skupiny dat párově – párové testy:	Párový t-test	Wilcoxonův test, znaménkový test
2 skupiny dat nepárově – dvouvýběrové testy:	Dvouvýběrový t-test	Mannův-Whitneyův test, mediánový test
Více skupin nepárově:	ANOVA	Kruskalův- Wallisův test

3. Kruskalův-Wallisův test

Co dělat, když nejsou splněny předpoklady u ANOVy?

- 1. Zkusit data transformovat** – např. logaritmická transformace by měla pomoci s normalizací rozdělení a stabilizací rozptylu u log-normálních dat.
- 2. Použít neparametrické testy** – např. Kruskalův-Wallisův test nevyžaduje předpoklad normality, pracuje stejně jako neparametrický Mannův-Whitneyův test.

Kruskalův-Wallisův test

- Neparametrická alternativa analýzy rozptylu (ANOVy).
- Testuje se, zda jsou srovnatelné distribuční funkce (obdobně jako u Mannova-Whitneyova testu).
- Hypotézy mají tvar: $H_0 : F_1(x) = F_2(x) = \dots = F_k(x)$
 $H_1 : \text{nejméně jedna } F_i \text{ je odlišná od ostatních}$
- Princip Kruskalova-Wallisova testu (podobný jako u Mannova-Whitneyova testu):
 1. Všechny hodnoty ze všech výběrů dohromady uspořádáme vzestupně podle velikosti \rightarrow každé hodnotě přiřadíme pořadí.
 2. Spočítáme součet pořadí hodnot u každého výběru.
 3. Na základě těchto dvou součtů vypočteme testovou statistiku.
- Tzn. za platnosti nulové hypotézy jsou spojená data dobře promíchaná a průměrná pořadí v jednotlivých souborech jsou podobná.
- Odlehlé hodnoty nejsou problém, protože pracujeme s pořadími.

Kruskalův-Wallisův test

- **Příklad:** Chceme srovnat, zda se liší MMSE skóre podle typu onemocnění.
- Tzn. hypotézy budou mít tvar: $H_0 : F_{AD}(x) = F_{MCI}(x) = F_{CN}(x)$
 $H_1 : \text{nejméně jedna } F_i \text{ je odlišná od ostatních}$
- **Postup:**
 1. Popisná sumarizace MMSE skóre podle typu onemocnění.
 2. Vykreslení histogramů MMSE skóre pro jednotlivé skupiny subjektů, abychom viděli, že není splněn předpoklad normálního rozdělení → proto použijeme neparametrický test.
 3. Aplikujeme statistický test.
 4. Nulovou hypotézu zamítneme nebo nezamítneme:
 $p < 0,001 < 0,05 \rightarrow \text{zamítáme}$ nulovou hypotézu → **MMSE skóre je u pacientů s AD, MCI a u kontrol statisticky významně odlišné.**
 5. Post hoc test:

Kruskal-Wallis test: H (2, N= 833) =492.6004 p =0.000				
Depend.:	CN	MCI	AD	
MMSE	R:651.95	R:419.32	R:137.92	
CN		0.00	0.00	
MCI	0.00		0.00	
AD	0.00	0.00		

→ rozdíl je mezi všemi skupinami

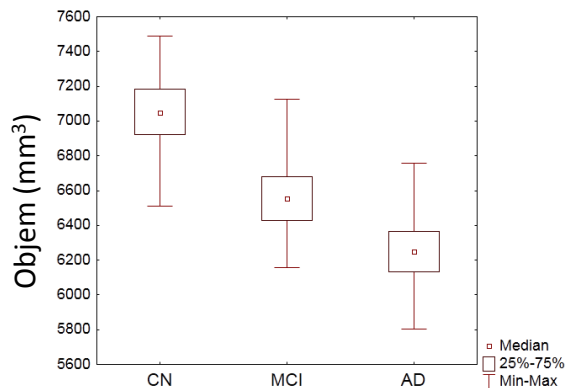
Úkol 2.

- **Zadání:** Zjistěte, zda se liší objem pěti mozkových struktur podle typu onemocnění (rozmyslete si, jaký test (či testy) byste použili).

Výsledky srovnání objemů mozkových podle typu onemocnění

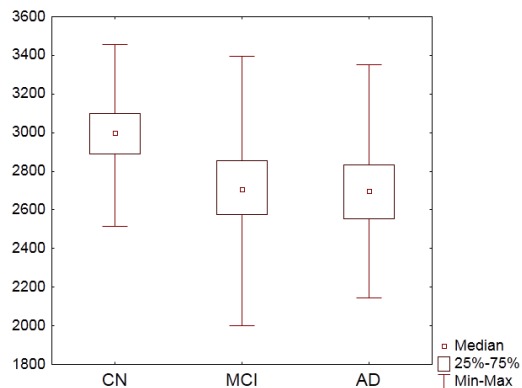
Hipokampus ($p < 0,001^*$)

* Statisticky významný rozdíl:
ADxMCI, ADxCN, MCIxCN

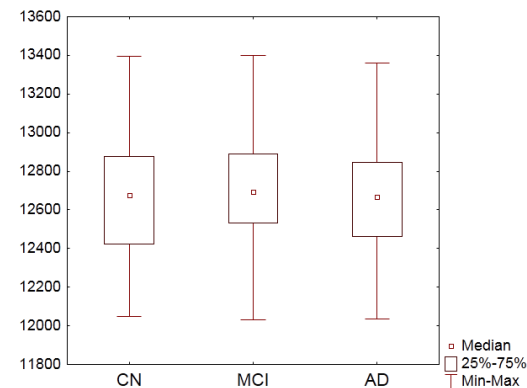


Amygdala ($p < 0,001^*$)

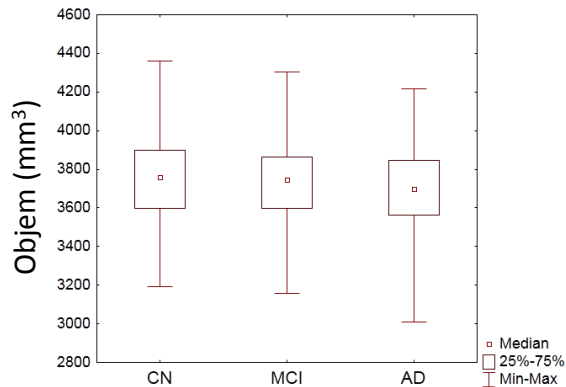
* Statisticky významný rozdíl:
ADxCN, MCIxCN



Thalamus ($p = 0,214$)

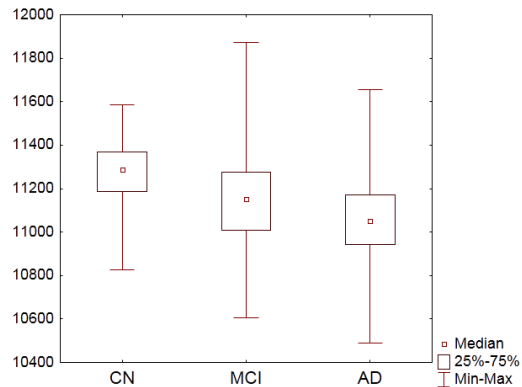


Pallidum ($p = 0,078$)



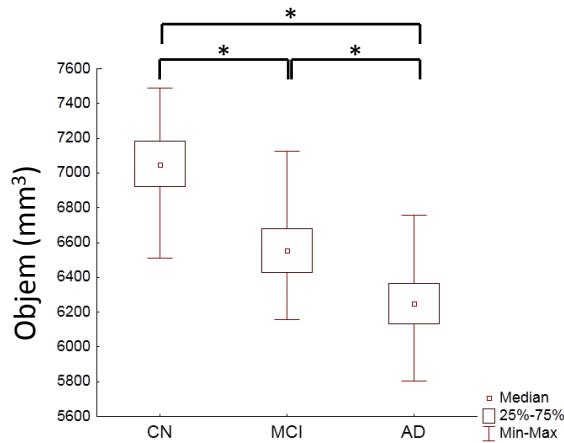
Putamen ($p < 0,001^*$)

* Statisticky významný rozdíl:
ADxMCI, ADxCN, MCIxCN

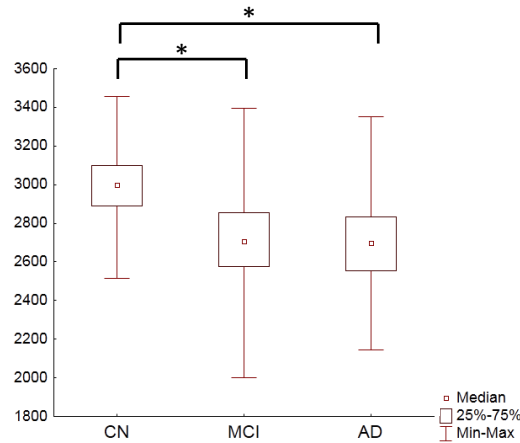


Výsledky srovnání objemů mozkových podle typu onemocnění – jiný způsob vyznačení lišících se skupin

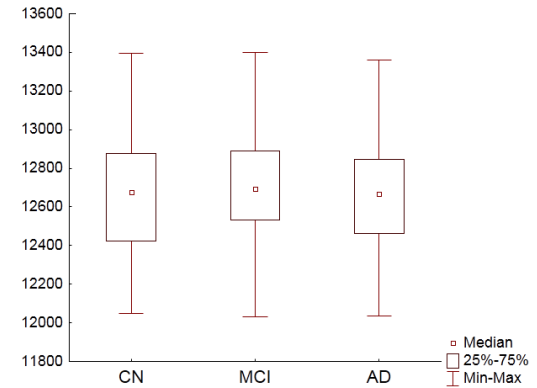
Hipokampus ($p < 0,001$)



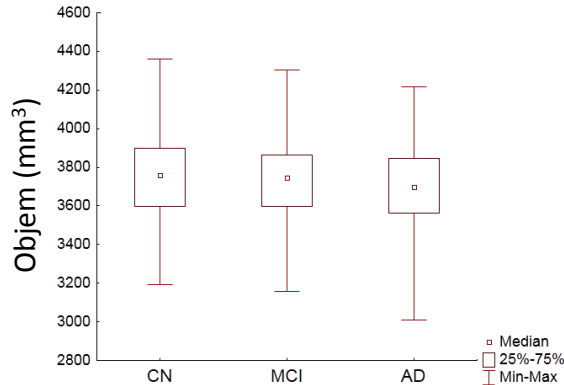
Amygdala ($p < 0,001$)



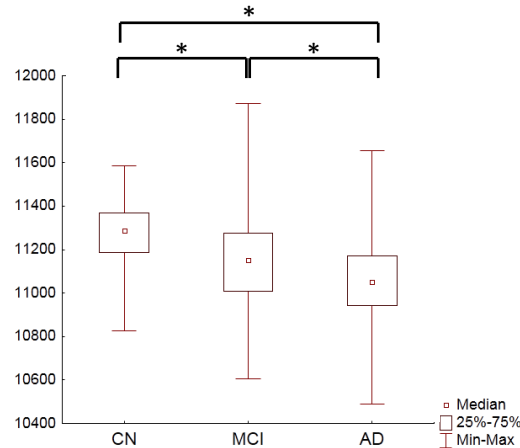
Thalamus ($p = 0,214$)



Pallidum ($p = 0,078$)



Putamen ($p < 0,001$)



Úkol 3.

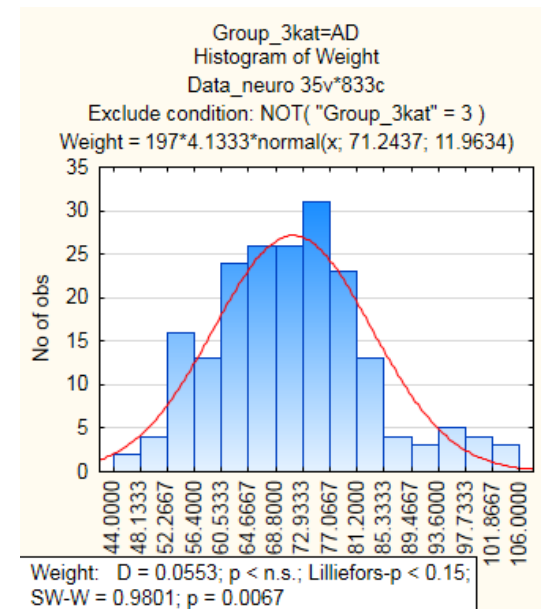
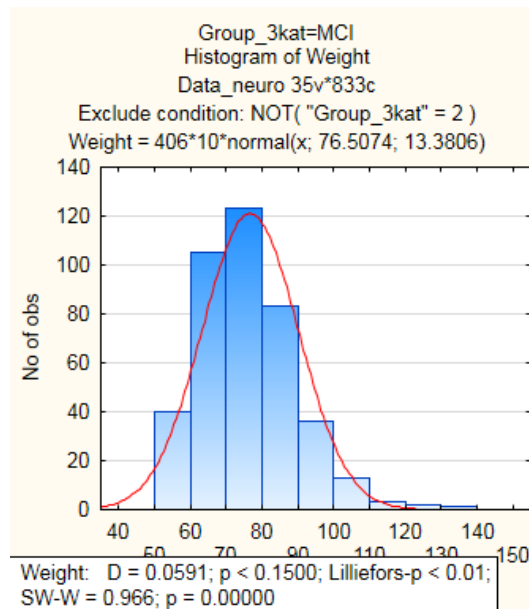
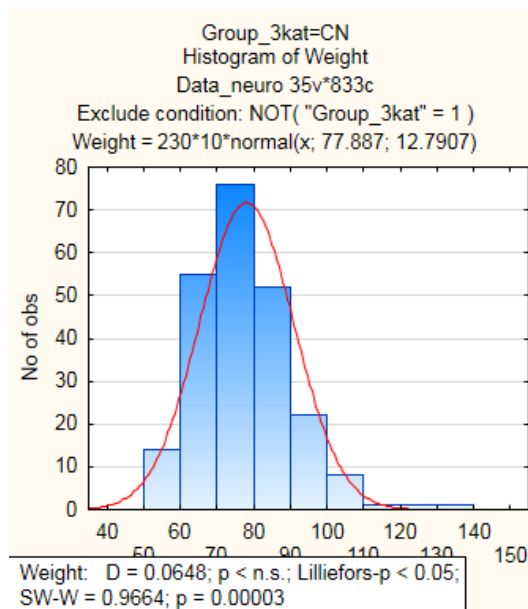
- **Zadání:** Zjistěte, zda se liší váha podle typu onemocnění. Pokud nejsou splněny předpoklady, zkuste váhu logaritmovat. Proveďte i popisnou sumarizaci váhy podle typu onemocnění včetně výpočtu intervalů spolehlivosti.

Úkol 3. – řešení

1. Provedeme popisnou sumarizaci:

Aggregate Results							
Descriptive Statistics (Data_neuro)							
Variable	Group_3kat	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.
Weight	CN	230	77.88696	76.00000	52.00000	135.0000	12.79073
Weight	MCI	406	76.50739	75.50000	52.00000	140.0000	13.38061
Weight	AD	197	71.24365	70.00000	44.00000	106.0000	11.96344

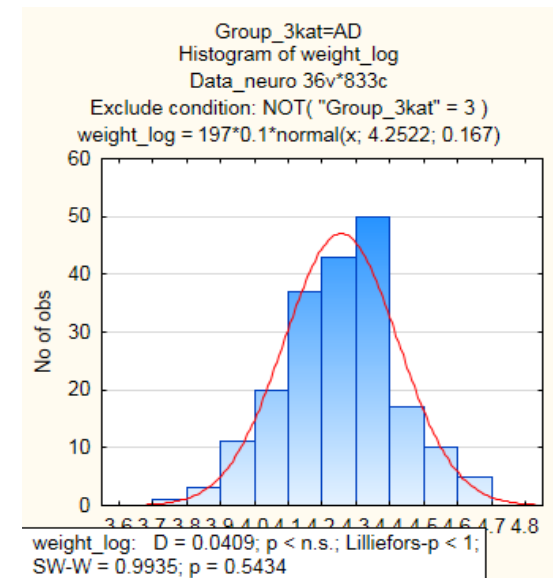
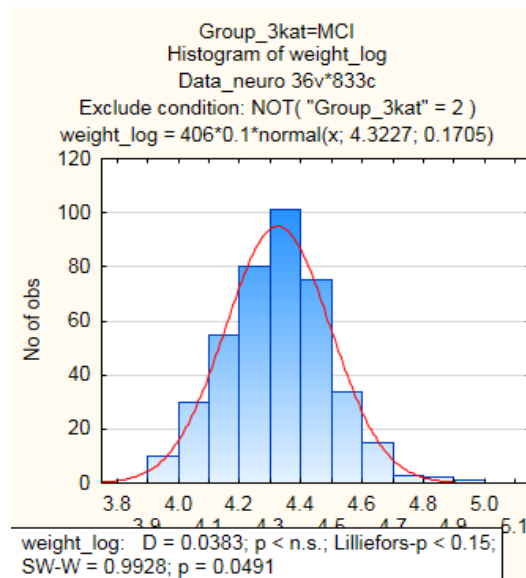
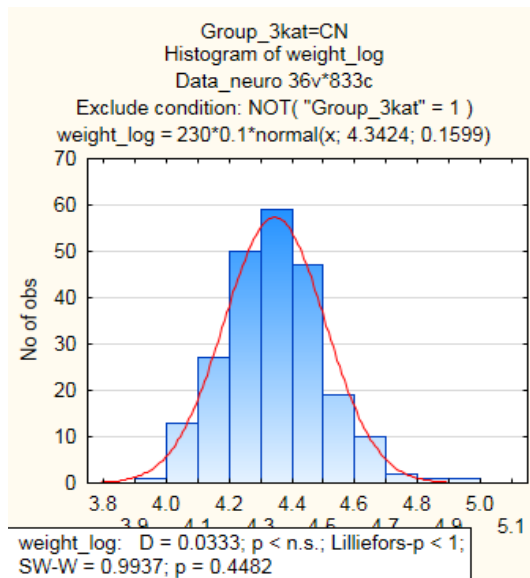
2. Ověření předpokladů (ověření normality je nutné i proto, abychom byli schopni říci, zda můžeme vypočítat intervaly spolehlivosti):



Úkol 3. – řešení

3. Protože data nemají normální rozdělení, zkusíme provést logaritmickou transformaci:
 Insert → Add Variables → Name: weight_log → Long name: =Log(Weight) nebo =Log10(Weight)

4. Ověření normality logaritmicky transformovaných dat:



Podle SW-W testu sice u MCI není splněn předpoklad normality, víme ale, že pro velká N je tento test až příliš přísný; podle histogramu však mají data normální rozdělení.

5. Ověření homogenity rozptylu u logaritmicky transformovaných dat:

Levene Test of Homogeneity of Variances (Data_neuro)								
Marked effects are significant at p < .05000								
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
weight_log	0.011865	2	0.005933	8.491216	830	0.010230	0.579900	0.560181

→ ANOVA

Úkol 3. – řešení

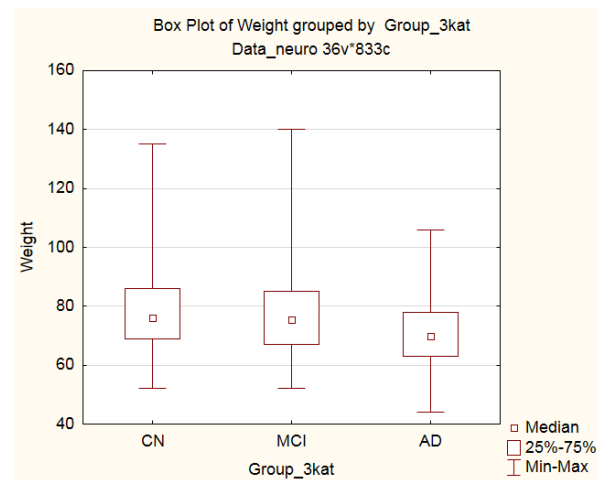
6. Protože byly splněny předpoklady, použijeme ANOVu pro zjištění, zda jsou mezi skupinami statisticky významné rozdíly ve váze (musíme použít logarimovanou váhu!):

Analysis of Variance (Data_neuro)								
Marked effects are significant at p < .05000								
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
weight_log	0.962059	2	0.481029	23.10202	830	0.027834	17.28223	0.000000

7. Mezi skupinami je statisticky významný rozdíl → musíme zjistit, mezi kterými skupinami ten rozdíl ve skutečnosti je:

Unequal N HSD; Variable: weight_log (Data_neuro)			
Marked differences are significant at p < .05000			
Group_3kat	{1}	{2}	{3}
	M=4.3424	M=4.3227	M=4.2522
CN {1}		0.416386	0.000022
MCI {2}	0.416386		0.000100
AD {3}	0.000022	0.000100	

Statisticky významné rozdíly u ADxMCI, ADxCN. Tzn. pacienti s AD mají statisticky významně nižší váhu než pacienti s MCI a zdravé kontroly.



Úkol 3. – řešení

8. Pokud chceme u popisné statistiky vypočítat intervaly spolehlivosti, nemůžeme je počítat na původních datech, protože neměla normální rozdělení, ale na datech po logaritmické transformaci:

Aggregate Results								
Descriptive Statistics (Data: neuro)								
Variable	Group_3kat	Valid N	Mean	Confidence -95.000%	Confidence 95.000%	Median	Minimum	Maximum
weight_log	CN	230	4.342363	4.321583	4.363144	4.330733	3.951244	4.905275
weight_log	MCI	406	4.322719	4.306083	4.339355	4.324111	3.951244	4.941642
weight_log	AD	197	4.252237	4.228768	4.275707	4.248495	3.784190	4.663439

9. Nyní bychom ale potřebovali transformovat tyto vypočtené hodnoty do původního měřítka, aby se nám výsledky lépe interpretovaly:

Zkopírujeme tabulku do Excelu (pomocí Copy with Headers) a provedeme exponenciální transformaci (pokud jsme ve STATISTICE použili funkci **Log**, použijeme v Excelu **=EXP(buňka)**; pokud **Log10**, použijeme v Excelu **=10^buňka**)

	N	Geometrický průměr	Dolní mez IS	Horní mez IS	Medián	Minimum	Maximum
CN	230	76,9	75,3	78,5	76,0	52,0	135,0
MCI	406	75,4	74,1	76,7	75,5	52,0	140,0
AD	197	70,3	68,6	71,9	70,0	44,0	106,0

Poznámka: po exponenciální transformaci průměru vypočteného na logaritmované váze dostáváme geometrický průměr)

Úkol 3. – srovnání popisné sumarizace

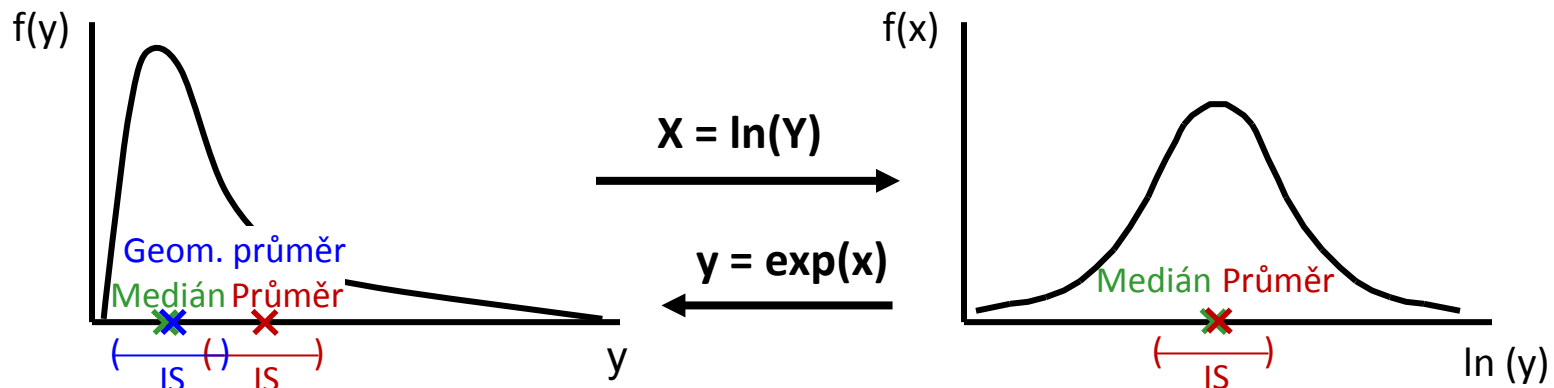
Popisná sumarizace původních hodnot váhy (sloupeček Weight):

Variable	Group_3kat	Valid N	Mean	Confidence -95.000%	Confidence 95.000%	Median	Minimum	Maximum
Weight	CN	230	77.88696	76.22515	79.54876	76.00000	52.00000	135.0000
Weight	MCI	406	76.50739	75.20194	77.81284	75.50000	52.00000	140.0000
Weight	AD	197	71.24365	69.56268	72.92463	70.00000	44.00000	106.0000

Popisná sumarizace hodnot zlogaritmované váhy po exponenciální transformaci:

	N	Geometrický průměr	Dolní mez IS	Horní mez IS	Medián	Minimum	Maximum
CN	230	76,9	75,3	78,5	76,0	52,0	135,0
MCI	406	75,4	74,1	76,7	75,5	52,0	140,0
AD	197	70,3	68,6	71,9	70,0	44,0	106,0

Je patrné, že medián, minimum i maximum jsou stejné. Avšak průměr a intervaly spolehlivosti vypočtené na původních datech jsou vyšší (tzn. nerepresentují dobře střed dat a jeho spolehlivost)!



4. Analýza rozptylu jako lineární model

Analýza rozptylu jako lineární model

- Analýza rozptylu pro jednu vysvětlující proměnnou (jednoduché třídění) lze zapsat jako lineární model:

$$Y_{ij} = \mu_i + e_{ij} = \mu + \alpha_i + e_{ij}$$

Populační průměr (arrow to μ)

i -tý efekt faktoru A (arrow to α_i)

Reziduum (arrow to e_{ij})

- Nulovou hypotézu pak lze vyjádřit jako: $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_k$
- Rozšířením tohoto zápisu můžeme definovat další modely ANOVA:** více faktorů, hodnocení interakcí, opakovaná měření na jednom subjektu.

Analýza rozptylu dvojného třídění

- Uvažujeme dvě vysvětlující proměnné zároveň.
- Zápis modelu:

$$Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$

↑ Populační průměr
 ↑ i -tý efekt faktoru A
 ← j -tý efekt faktoru B
 ← Reziduum

- Nulové hypotézy pak máme dvě: $H_{01} : \alpha_1 = \alpha_2 = \dots = \alpha_k$, $H_{02} : \beta_1 = \beta_2 = \dots = \beta_r$

Variabilita	Součet čtverců	Počet stupňů volnosti	Průměrný čtverec	F statistika	p-hodnota
Faktor A	S_A	$df_A = k - 1$	$MS_A = S_A / df_A$	F_A	p
Faktor B	S_B	$df_B = r - 1$	$MS_B = S_B / df_B$	F_B	p
Rezidua	S_e	$df_e = (k - 1)(r - 1)$	$MS_e = S_e / df_e$		
Celkem	S_T	$df_T = n - 1 = kr - 1$			

Analýza rozptylu dvojného třídění s interakcí

- Uvažujeme dvě vysvětlující proměnné a zároveň i jejich společné působení.

- Zápis modelu:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ij}$$

Diagrammatic annotations for the equation above:

- Blue arrow from μ to "Populační průměr"
- Blue arrow from α_i to " i -tý efekt faktoru A"
- Blue arrow from β_j to " j -tý efekt faktoru B"
- Red arrow from γ_{ij} to "Interakce"
- Blue arrow from e_{ij} to "Reziduum"

- Nulové hypotézy pak máme tři:

$$H_{01} : \gamma_{11} = \gamma_{12} = \dots = \gamma_{kr} \quad H_{02} : \alpha_1 = \alpha_2 = \dots = \alpha_k \quad H_{03} : \beta_1 = \beta_2 = \dots = \beta_r$$

Variabilita	Součet čtverců	Počet stupňů volnosti	Průměrný čtverec	F statistika	p -hodnota
Faktor A	S_A	$df_A = k - 1$	$MS_A = S_A / df_A$	F_A	p
Faktor B	S_B	$df_B = r - 1$	$MS_B = S_B / df_B$	F_B	p
Interakce AxB	S_{AB}	$df_{AB} = (k - 1)(r - 1)$	$MS_{AB} = S_{AB} / df_{AB}$	F_{AB}	p
Rezidua	S_e	$df_e = n - kr$	$MS_e = S_e / df_e$		
Celkem	S_T	$df_T = n - 1$			

Poděkování...

Příprava výukových materiálů předmětu „DSAN01 Analýza dat pro Neurovědy “ byla finančně podporována prostředky projektu FRVŠ č. 942/2013 „Inovace materiálů pro interaktivní výuku a samostudium předmětu Analýza dat pro Neurovědy“

