

Přednáška IX.

Analýza rozptylu (ANOVA)

- ➔ Princip a metodika výpočtu
- ➔ Předpoklady analýzy rozptylu a jejich ověření
- ➔ Rozbor rozdílů jednotlivých skupin – násobné testování hypotéz
- ➔ Analýza rozptylu jako lineární model



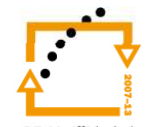
evropský
sociální
fond v ČR



EVROPSKÁ UNIE



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



OP Vzdělávání
pro konkurenceschopnost



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

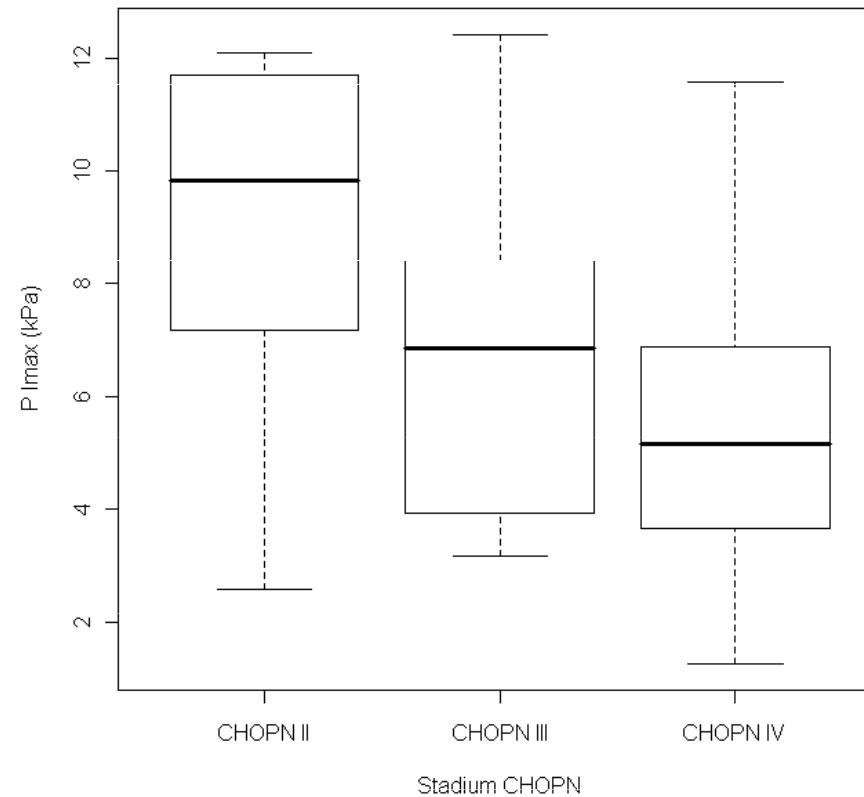
Opakování – parametrické a neparametrické testy

- ➡ Jmenujte příklad parametrického a neparametrického testu.
- ➡ Znáte jejich předpoklady?

1. Motivace

Příklad – CHOPN

- ➔ Sledujeme plicní funkce u pacientů s chronickou obstrukční plicní nemocí (CHOPN) ve stadiu II, III a IV. Zajímá nás, jestli se u pacientů v jednotlivých stadiích liší maximální inspirační tlak (P_{lmax}).



Příklad – CHOPN

➡ Jak můžeme pro CHOPN stadia II, III a IV ověřit rozdíl (resp. rovnost) v maximálním inspiračním tlaku ($P_{I\max}$)?

A. Můžeme použít vhodný test pro dva výběry (např. t -test) a otestovat, jak se liší stadium II od III, II od IV a III od IV – tedy provést 3 testy.

B. Musíme použít vhodný test pro více než dvě srovnávané skupiny.

➡ V čem je zásadní rozdíl mezi A a B?

Problém násobného testování hypotéz

➡ **Problém s možností A je v násobném testování hypotéz – pro připomenutí:**

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.**

Analýza rozptylu

➡ Lepší volbou je:

B. Musíme 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.

➡ Více než dvě skupiny mohou být dány přirozeně (např. sledujeme rozdíl mezi věkovými kategoriemi) nebo uměle (např. sledujeme rozdíl v účinnosti několika typů léčby).

2. Princip výpočtu

Náhodné výběry a hypotéza

➡ Máme k nezávislých realizací náhodného výběru rozsahu: n_1, n_2, \dots, n_k .

➡ Předpoklady: $Y_{1j} \sim N(\mu_1, \sigma^2)$

$$Y_{2j} \sim N(\mu_2, \sigma^2)$$

\vdots

$$Y_{kj} \sim N(\mu_k, \sigma^2)$$



Normalita hodnot všech k výběrů



Homogenita rozptylů všech k výběrů

➡ Hypotézy:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

H_1 : nejméně jedno μ_i je odlišné od ostatních

Příklady – hypotézy

1. Liší se účinnost dvou různých dávek léčiva XYZ od placebo?

Střední hodnota účinnosti placebo, XYZ v dávce 1 a XYZ v dávce 2: $\mu_P, \mu_{XYZ_1}, \mu_{XYZ_2}$

$$H_0 : \mu_P = \mu_{XYZ_1} = \mu_{XYZ_2}$$

H_1 : nejméně jedno μ je odlišné od ostatních

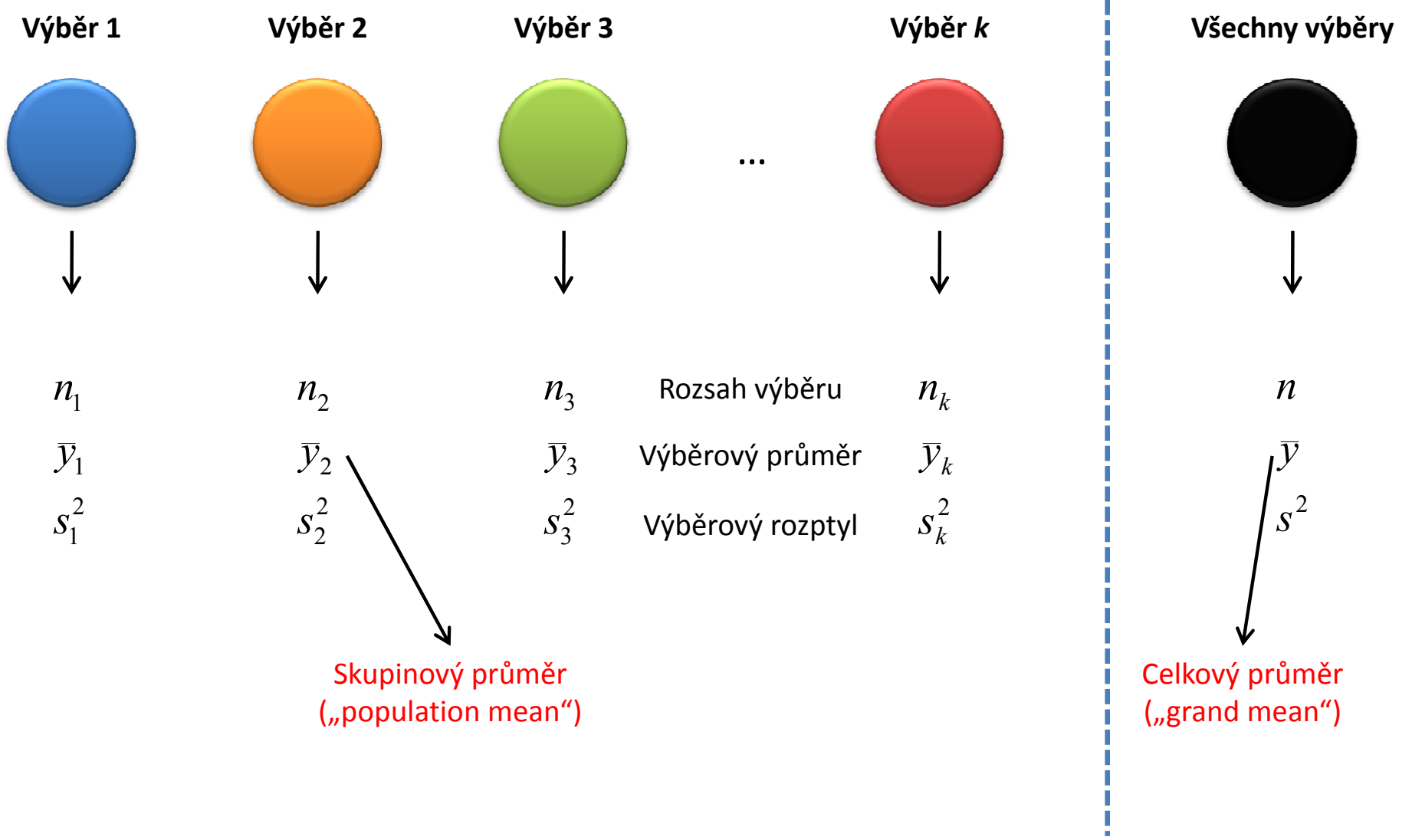
2. Liší se AML, ALL, CML a CLL v aktivitě vybraných genů?

Střední hodnota exprese genu g u AML, ALL, CML, CLL: $\theta_{AML}^g, \theta_{ALL}^g, \theta_{CML}^g, \theta_{CLL}^g$

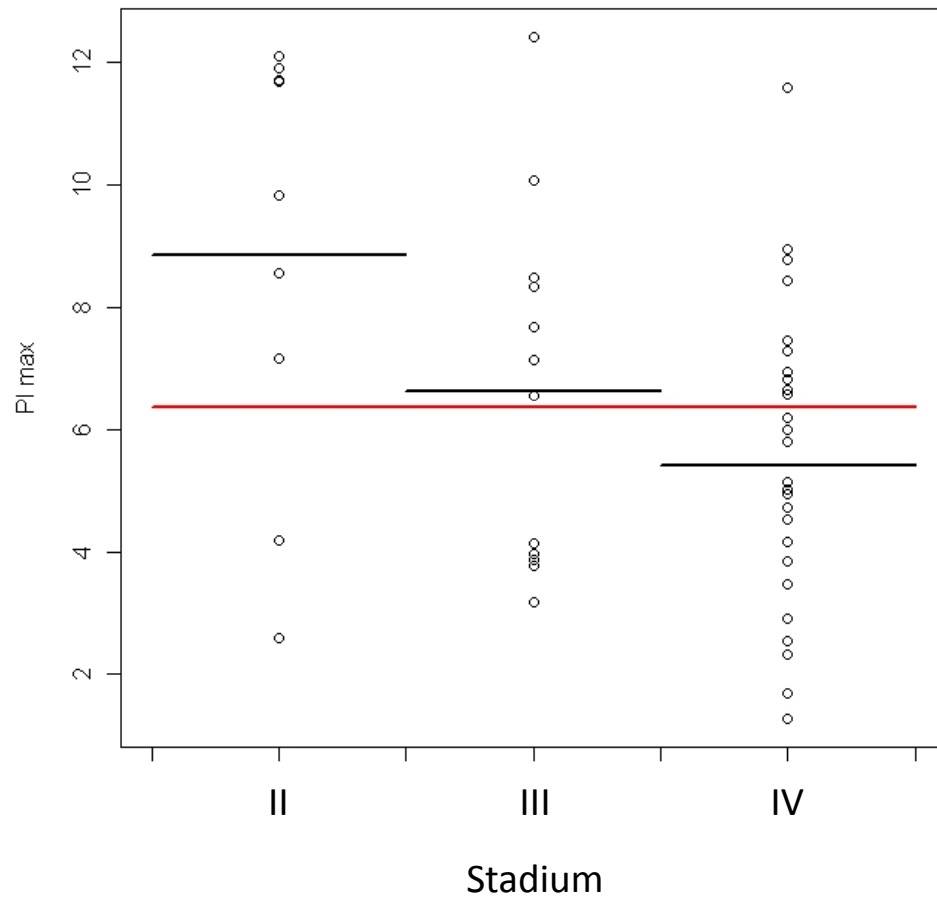
$$H_0 : \theta_{AML}^g = \theta_{ALL}^g = \theta_{CML}^g = \theta_{CLL}^g$$

H_1 : nejméně jedno θ^g je odlišné od ostatních

Pozorované hodnoty



Příklad – CHOPN



$$n_1 = 9$$

$$\bar{y}_1 = 8,9 \text{ kPa}$$

$$s_1 = 3,5 \text{ kPa}$$

$$n_2 = 12$$

$$\bar{y}_2 = 6,6 \text{ kPa}$$

$$s_2 = 2,9 \text{ kPa}$$

$$n_3 = 27$$

$$\bar{y}_3 = 5,4 \text{ kPa}$$

$$s_3 = 2,5 \text{ kPa}$$

Celkový průměr („grand mean“)

$$n = 48$$

$$\bar{y} = 6,4 \text{ kPa}$$

$$s = 3,0 \text{ kPa}$$

Značení

→ Součty:

$$Y_{i.} = \sum_{j=1}^{n_i} Y_{ij}$$

$$Y_{..} = \sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}$$

→ Průměry:

$$\bar{y}_{i.} = Y_{i.} / n_i$$

$$\bar{y}_{..} = Y_{..} / n$$

Skupinový průměr
(„population mean“)

Celkový průměr
(„grand mean“)

→ Celková variabilita v souboru:

$$S_T = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{y}_{..})^2$$

Stupně volnosti: $df_T = n - 1$

→ Variabilita v rámci skupin (reziduální součet čtverců):

$$S_e = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{y}_{i.})^2$$

Stupně volnosti: $df_e = n - k$

→ Variabilita mezi skupinami (příslušná sledovanému vlivu = proměnné):

$$S_A = \sum_{i=1}^k n_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

Stupně volnosti: $df_A = k - 1$

Vztahy mezi odhady variability

➔ Platí:

$$Y_{ij} - \bar{y}_{..} = (Y_{ij} - \bar{y}_{i.}) + (\bar{y}_{i.} - \bar{y}_{..})$$

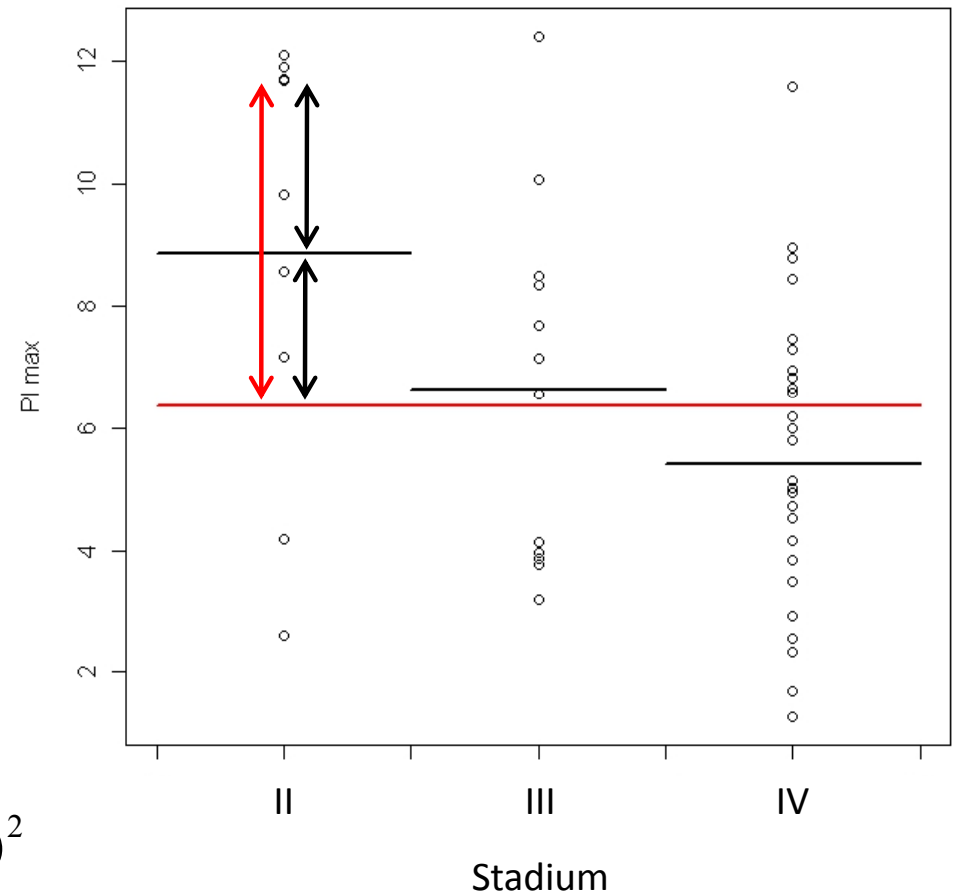
➔ Dále se dá ukázat, že platí:

$$S_T = S_e + S_A$$

➔ Tedy platí, že celková variabilita se dá rozložit na variabilitu v rámci skupin a variabilitu mezi skupinami:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{y}_{..})^2 =$$

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{y}_{i.})^2 + \sum_{i=1}^k n_i (\bar{y}_{i.} - \bar{y}_{..})^2$$



Umělý příklad

Léčba	Pozorovaná hodnota	Skupinový průměr	Skupinový průměr – celkový průměr	Pozorovaná hodnota – skupinový průměr	Pozorovaná hodnota – celkový průměr
1	10	12	-4	-2	-6
1	12	12	-4	0	-4
1	14	12	-4	2	-2
2	19	20	4	-1	3
2	20	20	4	0	4
2	21	20	4	1	5
3	14	16	0	-2	-2
3	16	16	0	0	0
3	18	16	0	2	2
	Celkový průměr = 16		Součet čtverců = 96	Součet čtverců = 18	Součet čtverců = 114
			Stupně volnosti = 2	Stupně volnosti = 6	Stupně volnosti = 8

Jak testuje t -test pro dva výběry?

➡ Nulová hypotéza:

$$H_0 : \mu_1 = \mu_2$$

➡ Testová statistika:

$$T = \frac{\bar{X} - \bar{Y}}{s_* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

Rozdíl (variabilita) mezi výběry

Variabilita uvnitř výběrů

➡

$$T = \frac{\text{Rozdíl (variabilita) mezi výběry}}{\text{Variabilita uvnitř výběrů}}$$

Princip analýzy rozptylu

➔ Princip analýzy rozptylu je stejný, tedy ANOVA srovnává pozorovanou variabilitu mezi výběry s pozorovanou variabilitou uvnitř výběrů. Na rozdíl od t -testu však pracuje s výběrovými rozptyly.

➔ Testová statistika analýzy rozptylu:

$$F = \frac{\text{Odhad rozptylu založený na výběrových průměrech}}{\text{Odhad rozptylu založený pozorovaných hodnotách}}$$

$$F = \frac{\frac{\sum_{i=1}^k n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}{k-1}}{\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{y}_{i\cdot})^2}{n-k}} = \frac{S_A / df_A}{S_e / df_e}$$

Za platnosti H_0 platí:

$$F \sim F(k-1, n-k)$$

Výsledek dle platnosti nulové hypotézy

- ➔ Za předpokladu rovnosti rozptylů jednotlivých výběrů představuje člen ve jmenovateli statistiky F výběrový odhad σ^2 .
- ➔ Za platnosti H_0 představuje i člen v čitateli statistiky F výběrový odhad σ^2 .
- ➔ **Platí-li nulová hypotéza, čítecel statistiky F (počítaný na základě výběrových průměrů) bude zhruba stejný jako její jmenovatel (počítaný na základě pozorovaných hodnot).**
- ➔ **Neplatí-li nulová hypotéza, čítecel statistiky F bude větší než jmenovatel.**
- ➔ **Samotné rozhodnutí o platnosti H_0 je tak založeno na srovnání průměrných čtverců S_A / df_A a S_e / df_e .**

Výsledek analýzy rozptylu

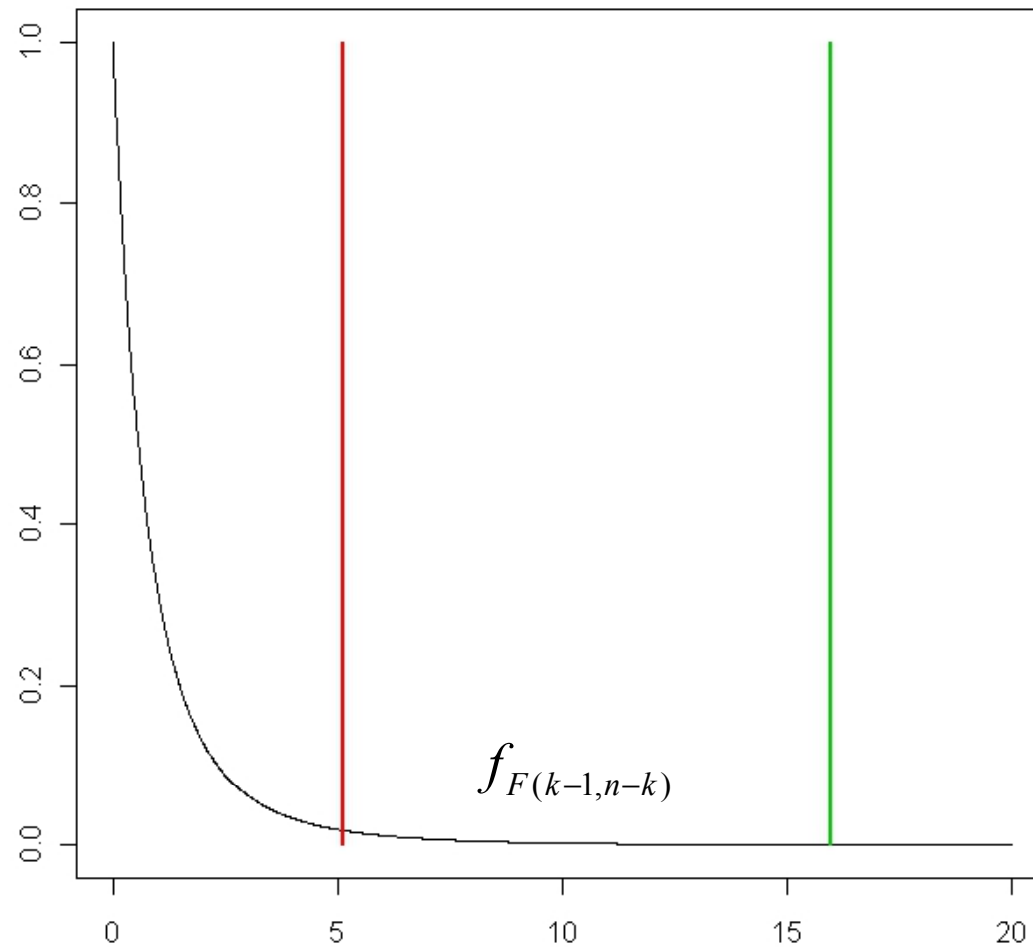
➔ Výsledné počty se standardně zaznamenávají do tzv. tabulky analýzy rozptylu:

Variabilita	Součet čtverců	Počet stupňů volnosti	Průměrný čtverec	F statistika	p-hodnota
Mezi skupinami	$S_A = 96$	$df_A = k - 1 = 2$	$MS_A = 48$	$F = 16$	0,004
Uvnitř skupin	$S_e = 18$	$df_e = n - k = 6$	$MS_e = 3$		
Celkem	$S_T = 114$	$df_T = n - 1 = 8$			

➔ Nulovou hypotézu zamítneme/nezamítneme buď na základě srovnání výsledné p-hodnoty se zvolenou hladinou významnosti testu α , nebo srovnáním výsledné F statistiky s kritickou hodnotou (kvantilem) rozdělení $F(k - 1, n - k)$ příslušnou zvolené hladině významnosti testu α .

Výsledek umělého příkladu

$$F_{1-\alpha}^{(k-1, n-k)} = F_{0,95}^{(2,6)} = 5,14 \quad F = 16$$



Na hladině významnosti $\alpha = 0,05$ zamítáme H_0 o rovnosti středních hodnot.

3. Předpoklady analýzy rozptylu a jejich ověření

Předpoklady analýzy rozptylu

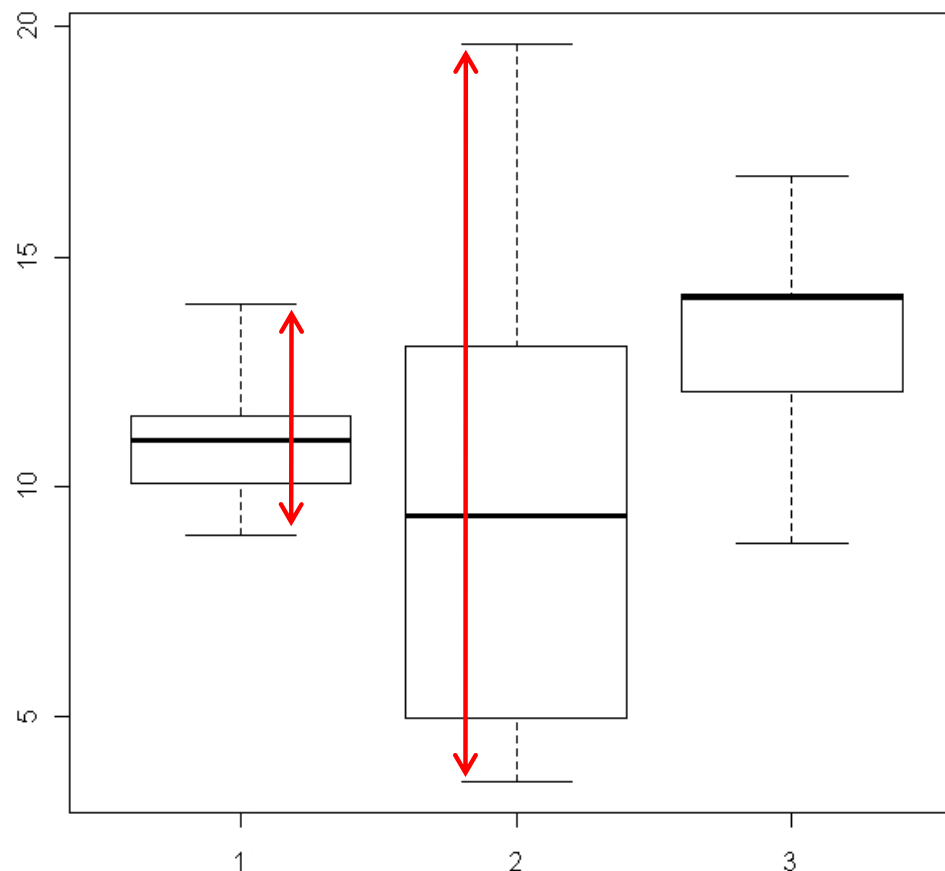
- ➔ **Nezávislost jednotlivých pozorování** – sice téměř automatický předpoklad, nicméně je třeba se nad ním alespoň zamyslet.
- ➔ **Normalita pozorovaných hodnot obou náhodných výběrů** – velmi silný předpoklad. **Nutno otestovat nebo alespoň graficky ověřit** (histogram, box plot).
- ➔ **Stejný rozptyl náhodné veličiny v obou srovnávaných skupinách** – také silný předpoklad. Opět **nutno otestovat nebo alespoň graficky ověřit** (histogram, box plot).

Testování shody rozptylů

➔ **Grafické ověření** –
histogram, box plot. ➔

➔ **Levenův test** – často
používaný, nevyžaduje
předpoklad normality
původních hodnot.

➔ **Bartlettův test** – velkou
nevýhodou je předpoklad
normality původních
hodnot.



Levenův test

➔ Jeho výhodou je, že **nevyžaduje předpoklad normality původních hodnot**.

➔ **Jedná se o analýzu rozptylu na hodnotách** $Z_{ij} = |Y_{ij} - \bar{y}_i|$

➔ Označme $\bar{z}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} Z_{ij}$ a $\bar{z}_{\cdot\cdot} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} Z_{ij}$

➔ Testová statistika:
$$W = \frac{\sum_{i=1}^k n_i (\bar{z}_{i\cdot} - \bar{z}_{\cdot\cdot})^2}{\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij} - \bar{z}_{i\cdot})^2}{n - k}}$$

Při rovnosti rozptylů opět platí:

$$F \sim F(k - 1, n - k)$$

➔ Používá se také jeho robustní varianta s použitím absolutních odchylek od mediánu místo od průměru: $Z_{ij} = |Y_{ij} - \tilde{y}_i|$

Příklad – Levenův test u CHOPN dat

- ➔ Sledujeme plicní funkce u pacientů s chronickou obstrukční plicní nemocí (CHOPN) ve stadiu II, III a IV.
- ➔ **Levenův test probíhá stejně jako jednoduchá ANOVA** – opět srovnáváme průměrné čtverce – reziduální a příslušné sledovaným faktorům.

Variabilita	Součet čtverců	Počet stupňů volnosti	Průměrný čtverec	F statistika	p-hodnota
Mezi skupinami	$S_A = 5,30$	$df_A = k - 1 = 2$	$MS_A = 2,65$	$F = 1,13$	0,331
Uvnitř skupin	$S_e = 105,35$	$df_e = n - k = 45$	$MS_e = 2,34$		
Celkem	$S_T = 110,65$	$df_T = n - 1 = 47$			

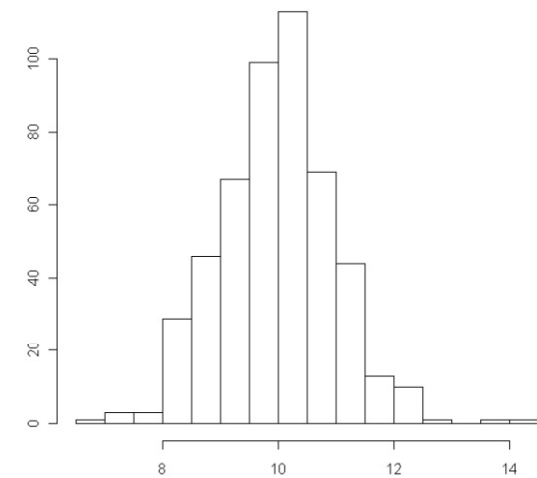
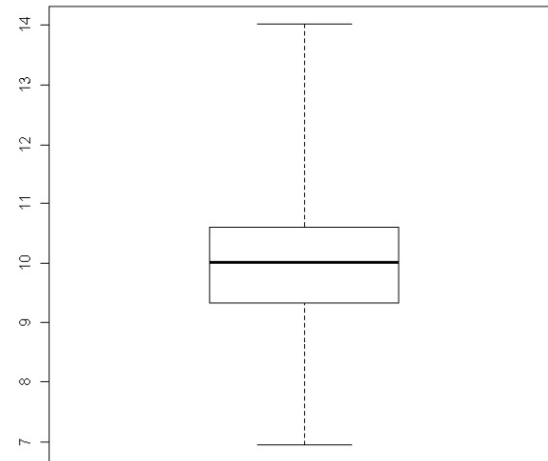
- ➔ Na hladině významnosti $\alpha = 0,05$ nezamítáme H_0 o rovnosti rozptylů.

Hodnocení normality dat

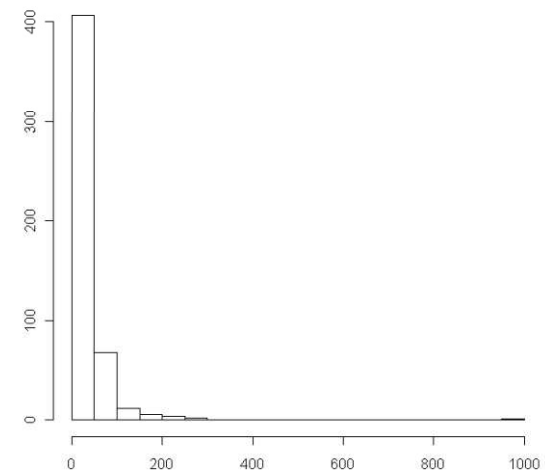
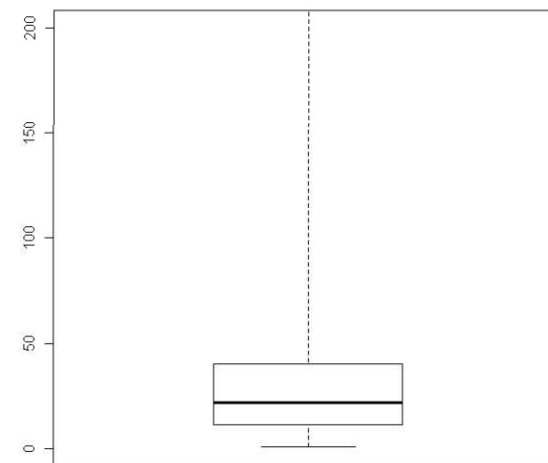
- ➡ **Hodnocení normality je klíčovým postupem v biostatistice. Testy nejsou vždy nejlepším nástrojem! Vždy je důležité se podívat i očima!**
- ➡ **Zamítnutí normality rozdělení** neznamena jenom výběr příslušného testu, **ALE může indikovat odlehlé a nelogické hodnoty v souboru dat.**
- ➡ Pokud o sledované 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.

Grafické metody – box plot a histogram

➡ Normální
rozdělení

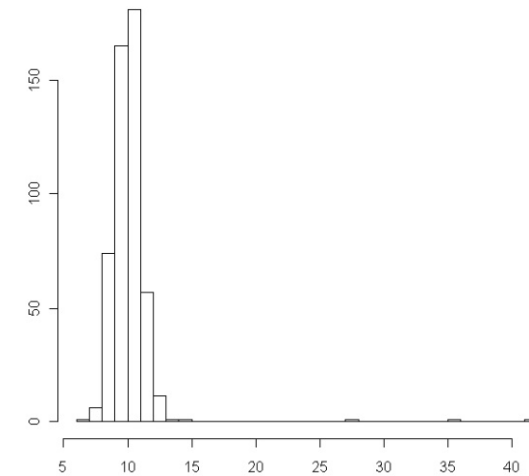
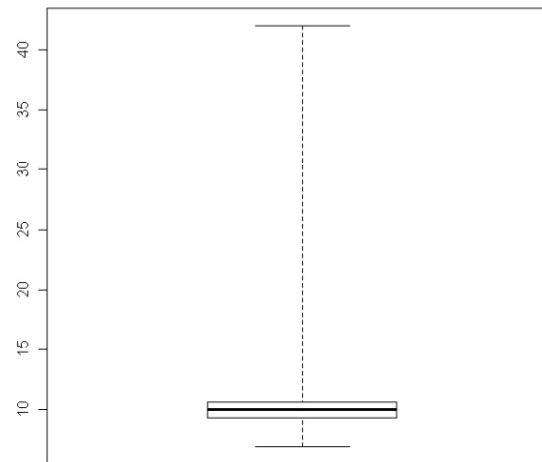


➡ Log-normální
rozdělení

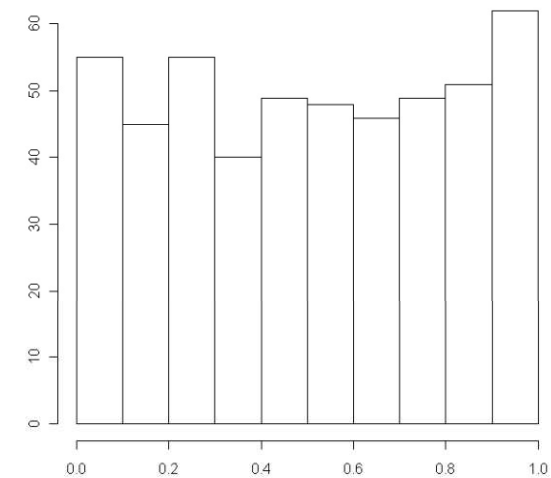
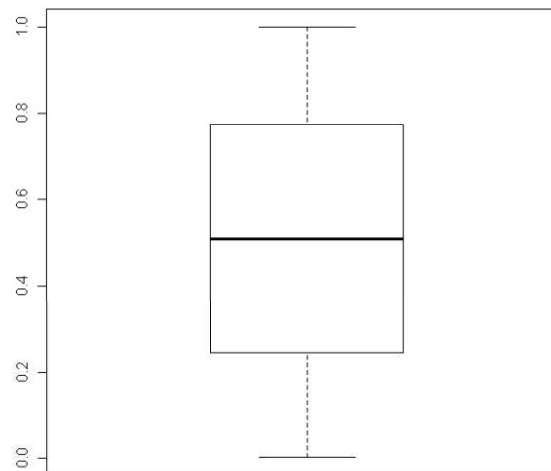


Grafické metody – box plot a histogram

➡ Normální
rozdělení
s odlehlými
hodnotami

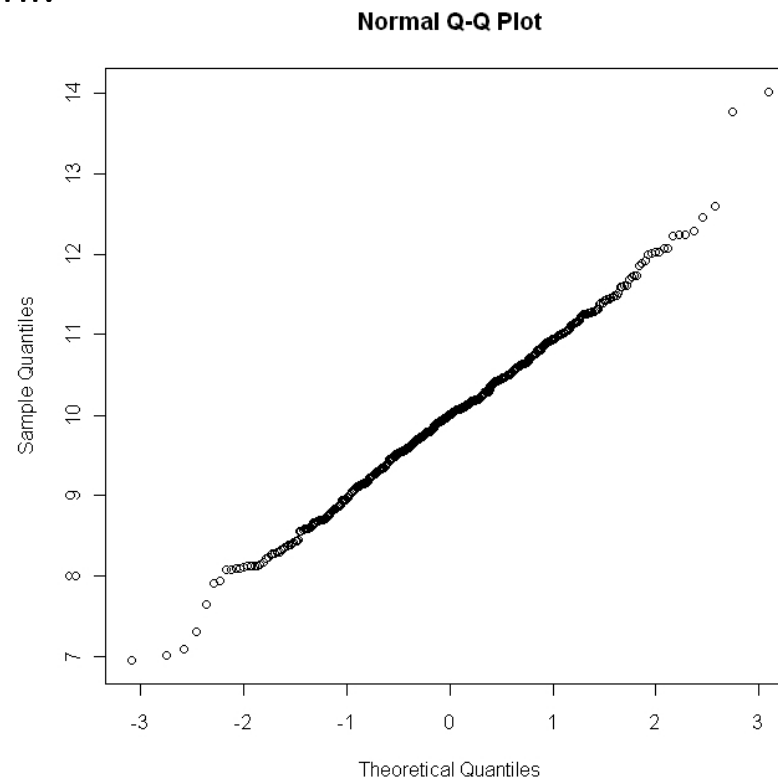


➡ Rovnoměrně
spojité
rozdělení



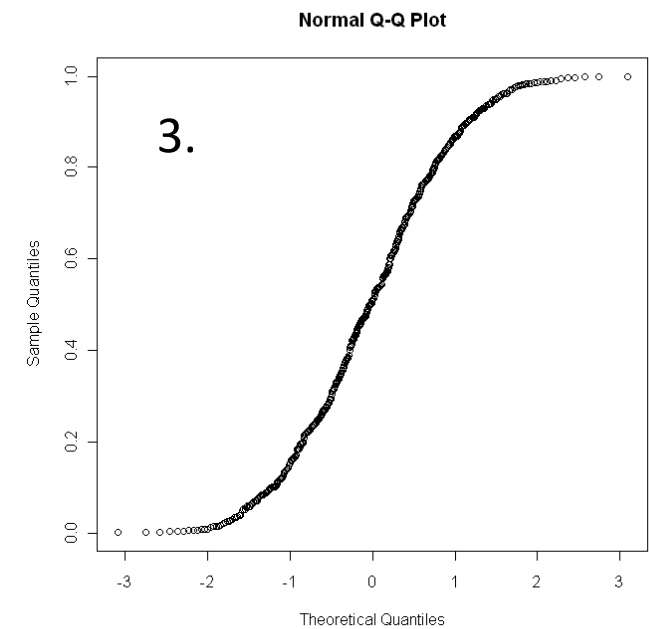
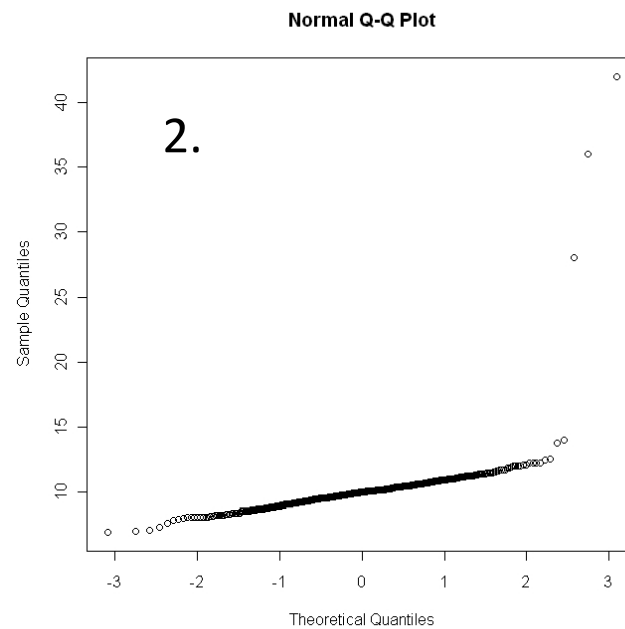
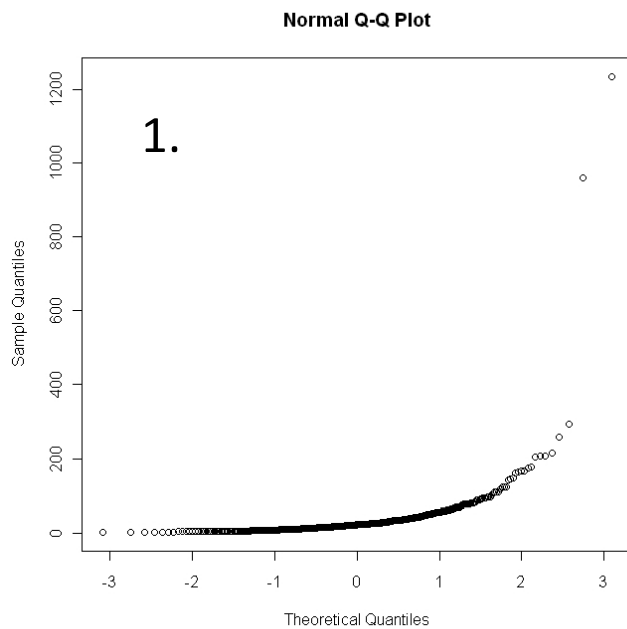
Grafické metody – Q-Q plot

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



Grafické metody – Q-Q plot

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



Testy pro ověření normality dat

- ➔ **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.

Příklad – Shapirův-Wilkův test u CHOPN dat

→ Sledujeme plicní funkce u pacientů s chronickou obstrukční plicní nemocí (CHOPN) ve stadiu II, III a IV.

→ Test pro všechna stadia:

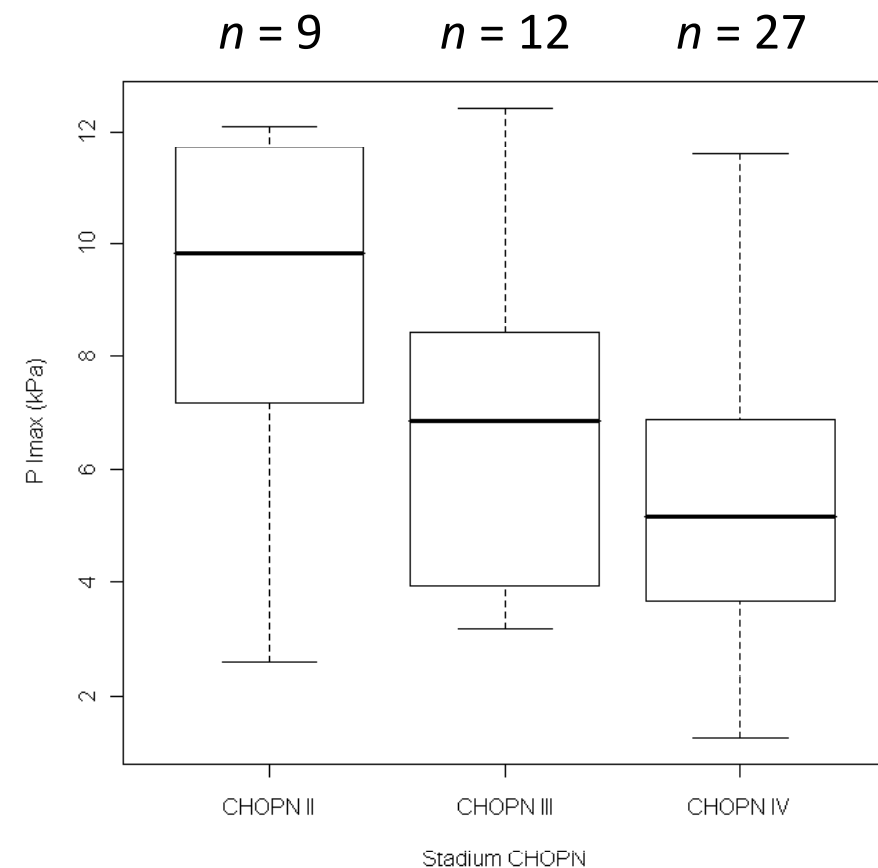
$p = 0,073$ (to nás nezajímá)

→ Stadium II: $p = 0,090$

→ Stadium III: $p = 0,247$

→ Stadium IV: $p = 0,815$

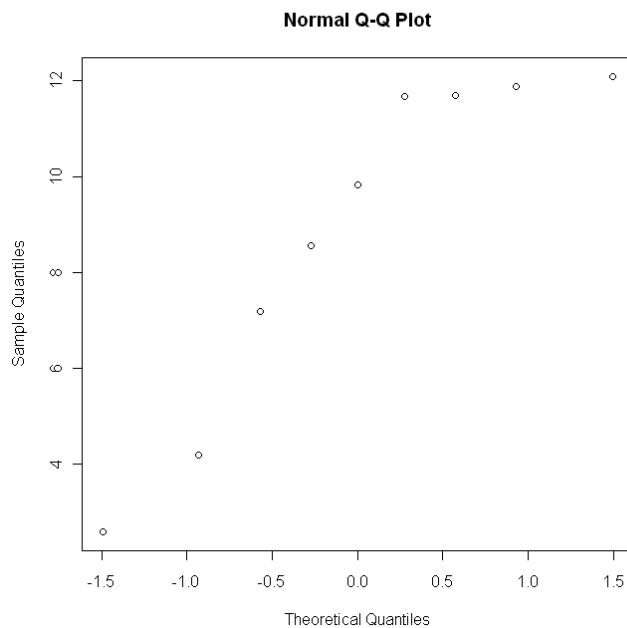
→ H_0 o normalitě dat nezamítáme na hladině $\alpha = 0,05$.



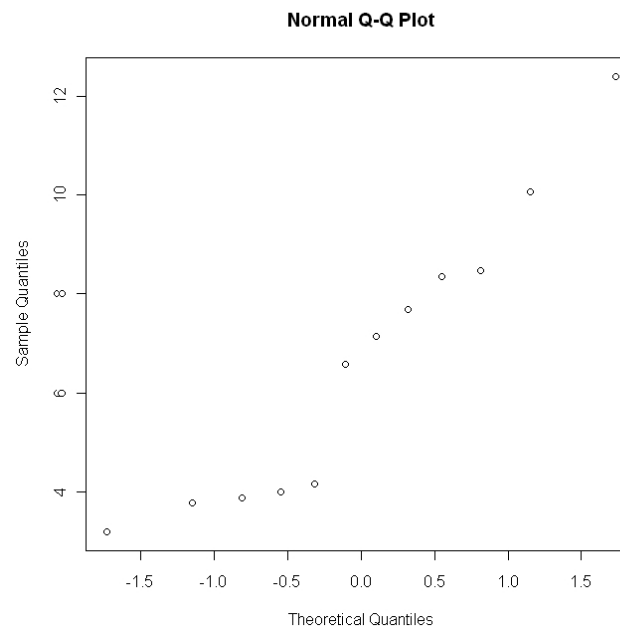
Příklad – Shapirův-Wilkův test u CHOPN dat

- ➡ Srovnáme výsledky S-W testu s Q-Q ploty pro jednotlivé kategorie.
- ➡ Vzhledem k malým velikostem souborů lze odchylky od normality dat tolerovat.

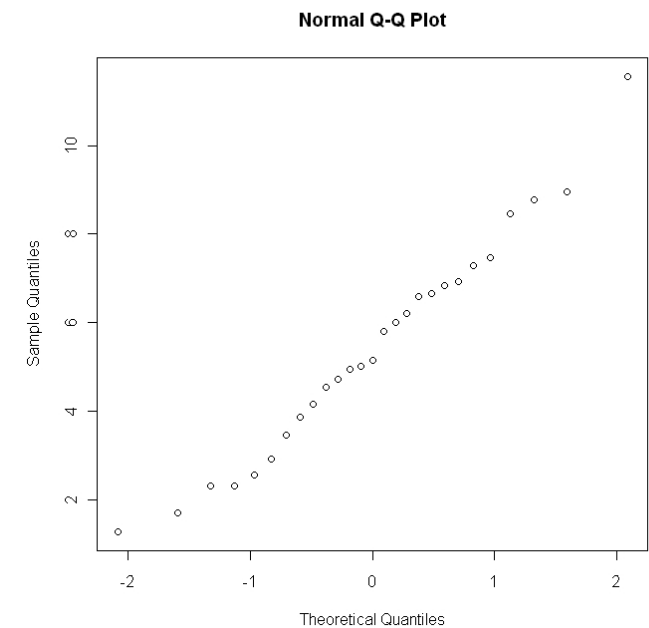
Stadium II ($n = 9$)
 $p = 0,090$



Stadium III ($n = 12$)
 $p = 0,247$



Stadium IV ($n = 27$)
 $p = 0,815$



Příklad – analýza rozptylu u CHOPN dat

- ➔ Liší se pacienti s CHOPN (stadium II, III, IV) v maximálním inspiračním tlaku ($P_{I\max}$)?
- ➔ Máme ověřenu homogenitu rozptylů i přibližnou normalitu dat → ANOVA.

Variabilita	Součet čtverců	Počet stupňů volnosti	Průměrný čtverec	F statistika	p-hodnota
Mezi skupinami	$S_A = 80,54$	$df_A = k - 1 = 2$	$MS_A = 40,27$	$F = 5,10$	0,010
Uvnitř skupin	$S_e = 355,50$	$df_e = n - k = 45$	$MS_e = 7,90$		
Celkem	$S_T = 436,04$	$df_T = n - 1 = 47$			

- ➔ Kritická hodnota pro $\alpha = 0,05$ $F_{(k-1, n-k)} = 3,20$.
- ➔ **Na hladině významnosti $\alpha = 0,05$ zamítáme H_0 o rovnosti středních hodnot.**

Co dělat, když nejsou splněny předpoklady?

➡ Máme dvě možnosti:

- 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

- ➔ Jedná se o zobecnění neparametrického Mannova – Whitneyho testu.
- ➔ Netestuje shodu parametrů, ale stejné distribuční funkce srovnávaných souborů (klíčový je zde předpoklad nezávislosti pozorovaných dat).

$$H_0 : F(x) = F(y) = F(z) = \dots$$

- ➔ Pro výpočet opět seřadíme všechna pozorování podle velikosti (jako by byly z jednoho vzorku) a přiřadíme jednotlivým hodnotám jejich pořadí.
- ➔ **Pointa Kruskalova – Wallisova testu:** za platnosti H_0 jsou spojená data dobře promíchaná a průměrná pořadí v jednotlivých souborech jsou podobná.

Kruskalův – Wallisův test

➔ Označme T_i součet pořadí v i -té skupině: $T_i = \sum_{j=1}^{n_i} R_{ij}$

➔ Počet skupin: k , Celkem pozorování: $n = n_1 + n_2 + \dots + n_k$.

➔ **Testová statistika:**
$$Q = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{T_i^2}{n_i} - 3(n+1)$$

➔ Nulovou hypotézu H_0 zamítáme na hladině významnosti, když je testová statistika větší nebo rovna kritické hodnotě chí-kvadrát rozdělení: $Q \geq \chi_{k-1}^2(\alpha)$

➔ Pro malé velikosti souboru je třeba srovnat statistiku Q s tabulkami pro Kruskalův-Wallisův test.

4. Násobné testování podskupin

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: **Bonferroniho procedura** - 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 pro velké m , tedy počet provedených testů.
- ➡ Pro analýzu rozptylu: **Tukeyho a Scheffého post hoc testy**.
- ➡ Pro neparametrický K-W test: **metoda dle Steela a Dwasse**.

Příklad – korekce u CHOPN dat

➔ ANOVA na hladině významnosti $\alpha = 0,05$ zamítla H_0 o rovnosti středních hodnot Plmax. Jaké skupiny se od sebe nejvíce liší?

➔ **Bonferroniho procedura**

Stadium	III	IV
II	0,398	0,009
III	-	0,571

➔ **Tukeyho post hoc test**

Stadium	III	IV
II	0,186	0,008
III	-	0,433

➔ **Scheffého post hoc test**

Stadium	III	IV
II	0,214	0,011
III	-	0,466

➔ Zde nám všechny tři analýzy vyšly stejně, ale obecně to neplatí!