UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

DIPLOMOVÁ PRÁCE

Logistická regrese se smíšenými efekty a její aplikace v medicíně

Katedra matematické analýzy a aplikací matematiky
Vedoucí diplomové práce: **doc. RNDr. Eva Fišerová, Ph.D.**
Vypracovala: **Bc. Veronika Hroudová**
Studijní program: Aplikovaná matematika (N0541A170026)
Studijní obor: Aplikovaná matematika
Forma studia: prezenční
Rok odevzdání: 2022

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Bc. Veronika Hroudová

Název práce: Logistická regrese se smíšenými efekty a její aplikace v medicíně

Typ práce: Diplomová práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: doc. RNDr. Eva Fišerová, Ph.D.

Rok obhajoby práce: 2022

Abstrakt: Diplomová práce se zabývá teorií a praktickým využitím logistické regrese s náhodným absolutním členem. V teoretické části jsou popsány metody odhadování a testování parametrů, výběru proměnných a hodnocení modelu nejprve logistické regrese a následně logistické regrese s náhodným absolutním členem. V praktické části jsou uvedeny výsledky analýzy dat týkajících se pacientů s nádorovým onemocněním mozku, tzv. meningeomem.

Klíčová slova: Zobecněný smíšený model, pevný efekt, náhodný efekt, logistická regrese, microRNA, meningeom

Počet stran: 77

Počet příloh: 4

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Bc. Veronika Hroudová

Title: Logistic regression with mixed effects and its application in medicine

Type of thesis: Master's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: doc. RNDr. Eva Fišerová, Ph.D.

The year of presentation: 2022

Abstract: The Master's thesis deals with the theory and practical use of logistic regression with random intercept. There are methods of estimating and testing parameters, selection of variables and evaluation of the model described in the theoretical part, firstly for logistic regression and then for logistic regression with random intercept. In the practical part, the results of data analysis of patients with brain tumor, meningioma, are presented.

Key words: Generalized mixed model, fixed effect, random effect, logistic regression, microRNA, meningioma

Number of pages: 77

Number of appendices: 4

Language: Czech

Prohlášení

Prohlašuji, že jsem diplomovou práci zpracovala samostatně pod vedením paní doc. RNDr. Evy Fišerové, Ph.D. a všechny použité zdroje jsem uvedla v seznamu literatury.

V Olomouci dne ....

podpis

Obsah

Poděkování

Ráda bych touto cestou poděkovala své vedoucí, paní doc. RNDr. Evě Fišerové, Ph.D., za její rady a čas, který mi věnovala. Dále bych chtěla poděkovat panu Mgr. Hanuši Slavíkovi, Ph.D. za poskytnutí dat a cenných poznámek ohledně kvantitativní PCR metody. V neposlední řadě chci poděkovat své rodině a přátelům za obrovskou podporu.

Úvod

Logistická regrese se smíšenými efekty slouží k modelování vztahu mezi binární náhodnou veličinou a nenáhodnými vysvětlujícími proměnnými. Své využití nalezne především v případě, kdy analyzujeme data, ve kterých se objevují skupiny závislých pozorování, např. data s opakovanými měřeními. Pro taková data nemůžeme využít prostředků klasické logistické regrese, jelikož bychom porušili předpoklad nezávislosti pozorování.

Cílem diplomové práce je seznámit čtenáře s problematikou logistické regrese s náhodným absolutním členem, získané znalosti aplikovat na data z medicínského prostředí a výsledky analýzy porovnat s výsledky studie, ze které data pochází.

Data byla pro účely této diplomové práce poskytnuta *Ústavem molekulární a translační medicíny* v Olomouci a výsledky jejich analýzy byly publikovány v článku *Identification of Meningioma Patients at High Risk of Tumor Recurrence Using MicroRNA Profiling* z roku 2020 [18]. Autoři článku pomocí analýzy přežití zkoumali, jaké faktory by mohly sloužit jako identifikátory pacientů s vyšším rizikem recidivy nádorového onemocnění mozku, tzv. meningeomu.

Diplomová práce je členěna do čtyř kapitol, teoretická část sestává z první a druhé kapitoly. V první kapitole se seznámíme s klasickou logistickou regresí, v druhé kapitole zavedeme logistický model s náhodným absolutním členem. Ve třetí kapitole čtenáři blíže představíme data, jejichž analýza je popsána ve čtvrté kapitole.

Kapitola 1

Vícenásobná logistická regrese

Regresní analýza je jedním ze základních a v praxi velmi často využívaných nástrojů matematické statistiky. Slouží k modelování vztahu mezi náhodnou vysvětlovanou (závisle) proměnnou a jednou či více nenáhodnými vysvětlujícími (nezávisle) proměnnými.

Logistická regrese je speciálním případem tzv. zobecněných lineárních modelů (GLM - *Generalized linear models*), kde uvažujeme binární závisle proměnnou. Binární proměnná nabývá pouze dvou hodnot a v praxi se s ní můžeme často setkat v roli závisle proměnné, např. pokud zkoumáme faktory, které ovlivňují, zda nastane, či nenastane recidiva onemocnění.

V rámci této kapitoly si nejprve zavedeme vícenásobný logistický regresní model. Následně ukážeme, jak lze interpretovat a odhadovat regresní parametry. Osvětlíme si též postup při testování významnosti regresních parametrů a hodnocení logistického modelu. Hlavními zdroji pro tuto kapitolu jsou [5], [6], [7], [8] a [17].

1.1. Vícenásobný logistický regresní model

Uvažujme náhodnou veličinu Y a sadu k nenáhodných vysvětlujících proměnných $x_1, x_2, \ldots, x_k$. Nechť se navíc náhodná veličina Y řídí alternativním rozdělením pravděpodobnosti.

Náhodná veličina řídící se alternativním rozdělením nabývá pouze dvou hod-

not. Ty jsou kódované pomocí čísel 1 (daný jev nastal) a 0 (daný jev nenastal). Pravděpodobnost nastání daného jevu se značí písmenem p a jde o jediný parametr tohoto rozdělení, který může nabývat pouze hodnot od 0 do 1. Píšeme tedy

$$P(Y = 1) = p, \qquad P(Y = 0) = 1 - p.$$

Dá se také velmi jednoduše odvodit, že pro střední hodnotu a rozptyl náhodné veličiny Y platí

$$\mathrm{E}(Y) = p, \qquad \mathrm{var}(Y) = p(1 - p).$$

Označme nyní ještě podmíněnou pravděpodobnost daného jevu při dané hodnotě $\boldsymbol{x} = (x_1, \ldots, x_k)'$ jako $\pi(\boldsymbol{x})$

$$P(Y = 1|\boldsymbol{x}) = \pi(\boldsymbol{x}), \qquad P(Y = 0|\boldsymbol{x}) = 1 - \pi(\boldsymbol{x}).$$

Na základě výše uvedených informací platí, že $\mathrm{E}(Y|\boldsymbol{x}) = \pi(\boldsymbol{x})$.

Klasický lineární regresní model, tedy model, ve kterém na rozdíl od logistického modelu uvažujeme kvantitativní závisle proměnnou, často píšeme ve tvaru

$$\mathrm{E}(Y|\boldsymbol{x}) = f(\boldsymbol{x}, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k, \tag{1.1}$$

kde $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$ je vektor neznámých regresních parametrů. Modelujeme tedy podmíněnou střední hodnotu náhodné veličiny Y pro dané hodnoty $\boldsymbol{x}$. V případě binární závisle proměnné Y je ale podmíněná střední hodnota $\mathrm{E}(Y|\boldsymbol{x}) = \pi(\boldsymbol{x})$ omezena pouze na hodnoty od 0 do 1 a regresní funkce $f(\boldsymbol{x}, \boldsymbol{\beta})$ ze vztahu (1.1) nám splnění tohoto omezení nezaručuje.

Proto je pro modelování střední hodnoty takové proměnné preferováno použití logistické regrese, která je založená na tzv. logitové transformaci funkce $\pi(\boldsymbol{x})$ dané jako

$$g(\boldsymbol{x}) = \ln\left(\frac{\pi(\boldsymbol{x})}{1 - \pi(\boldsymbol{x})}\right). \tag{1.2}$$

Za předpokladu, že podmíněná pravděpodobnost $\pi(\boldsymbol{x})$ nabývá pouze hodnot z otevřeného intervalu $(0, 1)$, nabývá zlomek $\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}$ pouze kladných hodnot. Přirozený logaritmus tohoto zlomku je tudíž definovaný a nabývá všech reálných

hodnot. Můžeme tedy říct, že funkce $g(\boldsymbol{x})$ je neomezená. Vícenásobný logistický regresní model pro n pozorování můžeme psát ve tvaru

$$g(\boldsymbol{x}_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}, \quad i = 1, 2, \ldots, n. \tag{1.3}$$

Pomocí několika úprav si z rovnice (1.2) můžeme vyjádřit $\pi(\boldsymbol{x})$

$$\begin{aligned} \exp[g(\boldsymbol{x})] &= \frac{\pi(\boldsymbol{x})}{1 - \pi(\boldsymbol{x})} \\ \exp[g(\boldsymbol{x})] - \pi(\boldsymbol{x}) \exp[g(\boldsymbol{x})] &= \pi(\boldsymbol{x}) \\ \exp[g(\boldsymbol{x})] &= \pi(\boldsymbol{x})(1 + \exp[g(\boldsymbol{x})]) \\ \pi(\boldsymbol{x}) &= \frac{\exp[g(\boldsymbol{x})]}{1 + \exp[g(\boldsymbol{x})]}. \end{aligned}$$

Po dosazení vztahu (1.3) získáme

$$\pi(\boldsymbol{x}_i) = \frac{\exp[g(\boldsymbol{x}_i)]}{1 + \exp[g(\boldsymbol{x}_i)]} = \frac{\exp[\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}]}{1 + \exp[\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}]}. \tag{1.4}$$

Podmíněná střední hodnota náhodné veličiny Y je tedy vyjádřena jako nelineární funkce k vysvětlujících proměnných. Funkce

$$f(x) = \frac{1}{1 + \exp(-x)} = \frac{\exp(x)}{1 + \exp(x)}$$

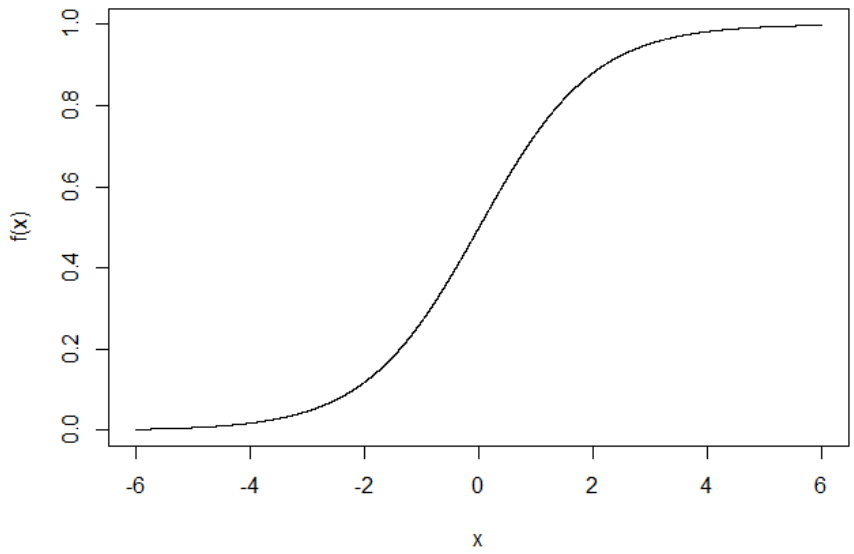
se nazývá logistická funkce (odtud název logistická regrese). Její graf je na Obrázku 1.1 a je často označován jako s-křivka či sigmoida.

Podotkněme, že logitová transformace není jedinou možností, jak transformovat binární data pro účely regresní analýzy. V kontextu zobecněných lineárních modelů je celá řada jiných tzv. spojovacích funkcí (*link functions*), které určují vztah mezi závisle proměnnou a regresní funkcí.

1.2. Interpretace regresních parametrů

Zadefinujme nejprve obecně pojem šance (*odds*) jevu A za podmínky jevu B jako

$$Odds(A|B) = \frac{P(A|B)}{P(A^C|B)} = \frac{P(A|B)}{1 - P(A|B)}$$



Obrázek 1.1: Graf logistické funkce

a poměr šancí (*odds ratio*) jako

$$OR = \frac{Odds(A|B)}{Odds(A|B^C)}.$$

Právě poměr šancí nám totiž velmi usnadní interpretaci regresních parametrů.

Mějme vícenásobný logistický regresní model s kvantitativními vysvětlujícími proměnnými $x_1, \ldots, x_k$

$$\ln\left(\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k.$$

Můžeme si všimnout, že logitová transformace podmíněné pravděpodobnosti $\pi(\boldsymbol{x})$ není nic jiného, než logaritmus šance jevu $Y = 1$ při daných hodnotách $\boldsymbol{x}$, tedy

$$\ln\left(Odds(Y = 1|\boldsymbol{x})\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k. \tag{1.5}$$

Přepišme si ještě vztah (1.5) jako

$$\ln\left(Odds(Y = 1|\boldsymbol{x})\right) = \beta_j x_j + C,$$

kde C je za předpokladu fixace proměnných $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k$ na neměnné hodnotě konstanta

$$C = \beta_0 + \beta_1 x_1 + \cdots + \beta_{j-1} x_{j-1} + \beta_{j+1} x_{j+1} + \cdots + \beta_k x_k.$$

Pokud si napíšeme poměr šancí, kde budeme v čitateli uvažovat šanci jevu $Y = 1$, když zvýšíme proměnnou x_j o jednotku, a ve jmenovateli šanci jevu $Y = 1$ při nezměněné hodnotě proměnné x_j, dostaneme

$$\begin{aligned} OR_j = \frac{Odds(Y = 1|x_1, \ldots, x_j + 1, \ldots, x_k)}{Odds(Y = 1|x_1, \ldots, x_j, \ldots, x_k)} &= \frac{\exp(\beta_j(x_j + 1) + C)}{\exp(\beta_j x_j + C)} \\ &= \exp(\beta_j x_j + \beta_j + C - \beta_j x_j - C) \\ &= \exp(\beta_j). \end{aligned}$$

Odtud tedy platí, že $\ln(OR_j) = \beta_j$. Místo přímé interpretace parametru β_j je proto jednodušší interpretovat spíše $\exp(\beta_j)$, a to jako násobnou změnu šance jevu $Y = 1$ při zvýšení proměnné x_j o jednotku a fixaci ostatních proměnných. Speciálně $\exp(\beta_0)$ pak interpretujeme jako šanci jevu $Y = 1$ při nulových hodnotách všech vysvětlujících proměnných.

Výše uvedená interpretace se týká pouze kvantitativních vysvětlujících proměnných. V praxi se ale často setkáváme i s kvalitativními vysvětlujícími proměnnými, typicky např. s pohlavím či věkovou skupinou. Kvalitativní vysvětlující proměnné zahrnujeme do modelu logistické regrese pomocí tzv. umělých proměnných (*dummy variables*), stejně jako u klasického regresního modelu. Umělé proměnné nabývají pouze hodnot 0 nebo 1 a kódují jednotlivé kategorie kvalitativní proměnné.

Ukažme si to na jednoduchém příkladu s pohlavím jako jedinou vysvětlující proměnnou. Zavedeme umělou proměnnou x následovně

$$x = \begin{cases} 1 & \ldots \text{muž}, \\ 0 & \ldots \text{žena}. \end{cases}$$

Regresní model je ve tvaru

$$\ln\left(Odds(Y = 1|\boldsymbol{x})\right) = \beta_0 + \beta_1 x,$$

což můžeme ekvivalentně zaspat jako

$$\ln\left(Odds(Y=1|\text{muž})\right) = \beta_0 + \beta_1,$$
$$\ln\left(Odds(Y=1|\text{žena})\right) = \beta_0.$$

Poměr šancí jevu $Y = 1$ pro ženy a pro muže píšeme jako

$$OR = \frac{Odds(Y=1|\text{muž})}{Odds(Y=1|\text{žena})} = \frac{\exp(\beta_0+\beta_1)}{\exp(\beta_0)} = \exp(\beta_1),$$

takže hodnota $\exp(\beta_1)$ představuje, kolikrát je větší šance jevu $Y = 1$ pro muže než pro ženy. Hodnota $\exp(\beta_0)$ je pak šance jevu $Y = 1$ pro ženy. Kategorii žen v tomto případě označujeme jako referenční kategorii.

Obdobně si poradíme i s modely s vysvětlujícími kvalitativními proměnnými s více než dvěma kategoriemi. Jednu z kategorií zvolíme jako referenční a ostatní budeme kódovat pomocí umělých proměnných. Pokud tedy máme proměnnou s r kategoriemi, k jejímu zahrnutí do modelu budeme potřebovat $r - 1$ umělých proměnných, tudíž i regresních parametrů.

1.3. Odhady regresních parametrů

Naším cílem nyní bude odhadnout regresní parametry $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$. V případě logistické regrese se k získání odhadů regresních parametrů používá metoda maximální věrohodnosti. Metoda maximální věrohodnosti, v angličtině *Maximum likelihood estimation* (MLE), je jednou z nejuniverzálnějších metod pro odhadování parametrů. Princip této metody spočívá v hledání takových parametrů, které maximalizují věrohodnost (*likelihood*), že naším modelem popsaný proces produkuje data, která jsme napozorovali. Velmi zjednodušeně řečeno hledáme parametry, které nejlépe odpovídají námi napozorovaným datům.

Mějme tedy logistický regresní model (1.3) a předpokládejme, že náhodná veličina Y má alternativní rozdělení. Věrohodnost i-té napozorované hodnoty y_i při daných hodnotách $\boldsymbol{x}_i$ je pro $i \in \{1, 2, \ldots, n\}$

$$\pi(\boldsymbol{x}_i)^{y_i}[1 - \pi(\boldsymbol{x}_i)]^{(1-y_i)}.$$

Věrohodnostní funkce $L(\boldsymbol{\beta})$ je dána jako sdružená pravděpodobnostní funkce všech napozorovaných hodnot. Za předpokladu, že data pocházejí z náhodného výběru, je tedy ve tvaru

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi(\boldsymbol{x}_i)^{y_i}[1 - \pi(\boldsymbol{x}_i)]^{(1-y_i)}.$$

Pro usnadnění následné maximalizace je často výhodné použít logaritmickou transformaci, čímž dostaneme tzv. logaritmickou věrohodnostní funkci $l(\boldsymbol{\beta})$. Ta v našem případě po několika málo úpravách vypadá následovně

$$l(\boldsymbol{\beta}) = \ln[L(\boldsymbol{\beta})] = \sum_{i=1}^{n} \Big\{ y_i \ln[\pi(\boldsymbol{x}_i)] + (1 - y_i) \ln[1 - \pi(\boldsymbol{x}_i)] \Big\}. \tag{1.6}$$

Nyní stačí pouze vyřešit úlohu maximalizace. Uvědomme si nejprve, že

$$\ln[\pi(\boldsymbol{x}_i)] = \ln\left[\frac{\exp[g(\boldsymbol{x}_i)]}{1 + \exp[g(\boldsymbol{x}_i)]}\right] = g(\boldsymbol{x}_i) - \ln\big[1 + \exp[g(\boldsymbol{x}_i)]\big]$$

a

$$\ln[1 - \pi(\boldsymbol{x}_i)] = \ln\left[\frac{1}{1 + \exp[g(\boldsymbol{x}_i)]}\right] = -\ln\big[1 + \exp[g(\boldsymbol{x}_i)]\big].$$

Logaritmickou věrohodnostní funkci nyní parciálně zderivujeme podle jednotlivých parametrů $\beta_0, \ldots, \beta_k$

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_0} &= \frac{\partial}{\partial \beta_0} \sum_{i=1}^{n} \Big\{ y_i g(\boldsymbol{x}_i) - y_i \ln\big[1 + \exp[g(\boldsymbol{x}_i)]\big] - (1 - y_i) \ln\big[1 + \exp[g(\boldsymbol{x}_i)]\big] \Big\} \\
&= \sum_{i=1}^{n} \left\{ y_i - y_i \frac{\exp[g(\boldsymbol{x}_i)]}{1 + \exp[g(\boldsymbol{x}_i)]} - (1 - y_i) \frac{\exp[g(\boldsymbol{x}_i)]}{1 + \exp[g(\boldsymbol{x}_i)]} \right\} \\
&= \sum_{i=1}^{n} \big[y_i - y_i \pi(\boldsymbol{x}_i) - (1 - y_i) \pi(\boldsymbol{x}_i) \big] \\
&= \sum_{i=1}^{n} \big[y_i - \pi(\boldsymbol{x}_i) \big]
\end{aligned}$$

a pro $j = 1, 2, \ldots, k$

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_j} = \frac{\partial}{\partial \beta_j} \sum_{i=1}^{n} \Big\{ y_i g(\boldsymbol{x}_i) - y_i \ln\big[1 + \exp[g(\boldsymbol{x}_i)]\big] - (1 - y_i) \ln\big[1 + \exp[g(\boldsymbol{x}_i)]\big] \Big\}$$

$$= \sum_{i=1}^{n} \left\{ y_i x_{ij} - y_i \frac{\exp[g(\boldsymbol{x}_i)]}{1+\exp[g(\boldsymbol{x}_i)]} x_{ij} - (1-y_i) \frac{\exp[g(\boldsymbol{x}_i)]}{1+\exp[g(\boldsymbol{x}_i)]} x_{ij} \right\}$$

$$= \sum_{i=1}^{n} \left[y_i x_{ij} - y_i \pi(\boldsymbol{x}_i) x_{ij} - (1-y_i) \pi(\boldsymbol{x}_i) x_{ij} \right]$$

$$= \sum_{i=1}^{n} x_{ij} \left[y_i - \pi(\boldsymbol{x}_i) \right].$$

Výsledné výrazy položíme rovny nule. Získáme tak soustavu $k+1$ rovnic

$$\sum_{i=1}^{n} [y_i - \pi(\boldsymbol{x}_i)] = 0, \qquad \sum_{i=1}^{n} x_{ij} [y_i - \pi(\boldsymbol{x}_i)] = 0, \quad j = 1, 2, \ldots, k.$$

Řešení této soustavy rovnic nelze až na speciální případy explicitně vyjádřit. Proto je třeba jej najít iterativně pomocí softwaru, k čemuž se nejčastěji používá Newtonův-Raphsonův algoritmus (viz [5]).

Výsledný odhad značíme $\hat{\boldsymbol{\beta}}$ a po jeho dosazení do vzorce (1.4) získáme odhad $\hat{\pi}(\boldsymbol{x}_i)$. Varianční matice vektoru $\hat{\boldsymbol{\beta}}$ je dána jako inverze Fisherovy informační matice $\boldsymbol{I}(\boldsymbol{\beta})$ v bodě $\hat{\boldsymbol{\beta}}$. Prvky matice $\boldsymbol{I}(\boldsymbol{\beta})$ spočítáme jako záporně vzaté druhé parciální derivace logaritmické věrohodnostní funkce $l(\boldsymbol{\beta})$,

$$\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_0^2} = \frac{\partial}{\partial \beta_0} \sum_{i=1}^{n} \left[y_i - \pi(\boldsymbol{x}_i) \right]$$

$$= -\sum_{i=1}^{n} \frac{\exp[g(\boldsymbol{x}_i)] \left[1+\exp[g(\boldsymbol{x}_i)] \right] - \left[\exp[g(\boldsymbol{x}_i)] \right]^2}{\left[1+\exp[g(\boldsymbol{x}_i)] \right]^2}$$

$$= -\sum_{i=1}^{n} \frac{\exp[g(\boldsymbol{x}_i)]}{1+\exp[g(\boldsymbol{x}_i)]} \frac{1}{1+\exp[g(\boldsymbol{x}_i)]}$$

$$= -\sum_{i=1}^{n} \pi(\boldsymbol{x}_i)[1-\pi(\boldsymbol{x}_i)],$$

podobně pro $j, l = 1, \ldots, k$

$$\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_0 \partial \beta_j} = \frac{\partial}{\partial \beta_j} \sum_{i=1}^{n} [y_i - \pi(\boldsymbol{x}_i)]$$

$$= -\sum_{i=1}^{n} x_{ij} \frac{\exp[g(\boldsymbol{x}_i)]}{1+\exp[g(\boldsymbol{x}_i)]} \frac{1}{1+\exp[g(\boldsymbol{x}_i)]}$$

$$= -\sum_{i=1}^{n} x_{ij} \pi(\boldsymbol{x}_i)[1-\pi(\boldsymbol{x}_i)]$$

a

$$\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_l} = \frac{\partial}{\partial \beta_l} \sum_{i=1}^{n} x_{ij}[y_i - \pi(\boldsymbol{x}_i)]$$

$$= -\sum_{i=1}^{n} x_{ij} x_{il} \frac{\exp[g(\boldsymbol{x}_i)]}{1+\exp[g(\boldsymbol{x}_i)]} \frac{1}{1+\exp[g(\boldsymbol{x}_i)]}$$

$$= -\sum_{i=1}^{n} x_{ij} x_{il} \pi(\boldsymbol{x}_i)[1-\pi(\boldsymbol{x}_i)].$$

Celkově a po dosazení odhadu $\hat{\boldsymbol{\beta}}$ tedy získáme, že

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = \boldsymbol{I}^{-1}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1},$$

kde matice $\boldsymbol{X}$ je matice plánu s rozměry $(k+1) \times n$ a $\boldsymbol{V}$ je diagonální matice s diagonálními prvky $\hat{\pi}(\boldsymbol{x}_i)[1-\hat{\pi}(\boldsymbol{x}_i)]$ o rozměrech $n \times n$, tj.

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix},$$

$$\boldsymbol{V} = \begin{pmatrix} \hat{\pi}(\boldsymbol{x}_1)[1-\hat{\pi}(\boldsymbol{x}_1)] & 0 & \dots & 0 \\ 0 & \hat{\pi}(\boldsymbol{x}_2)[1-\hat{\pi}(\boldsymbol{x}_2)] & \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \dots & 0 & \hat{\pi}(\boldsymbol{x}_n)[1-\hat{\pi}(\boldsymbol{x}_n)] \end{pmatrix}.$$

Odhad $\hat{\boldsymbol{\beta}}$ je asymptoticky nestranný a pro dostatečně velký počet pozorování má navíc přibližně normální rozdělení. Díky tomu dokážeme sestrojit přibližné

intervalové odhady jednotlivých regresních parametrů β_j, $j = 0, 1, \ldots, k$, se spolehlivostí $(1 - \alpha)$

$$\mathcal{I}_{1-\alpha}(\beta_j) = \left\langle \hat{\beta}_j - \mathrm{u}(1-\alpha/2)\sqrt{\mathrm{var}(\hat{\beta}_j)}, \quad \hat{\beta}_j + \mathrm{u}(1-\alpha/2)\sqrt{\mathrm{var}(\hat{\beta}_j)} \right\rangle, \qquad (1.7)$$

kde $\mathrm{u}(1-\alpha/2)$ značí $(1-\alpha/2)$-kvantil normovaného normálního rozdělení.

Jak jsme si ukázali v podkapitole 1.2, pro snadnější interpretaci regresních parametrů používáme poměry šancí OR_j, pro které platí, že $OR_j = \exp(\beta_j)$. Pro stanovení intervalových odhadů jednotlivých poměrů šancí tedy stačí pouze aplikovat exponenciální transformaci na intervalové odhady regresních parametrů ze vztahu (1.7)

$$\mathcal{I}_{1-\alpha}(OR_j) = \left\langle \exp\left(\hat{\beta}_j - \mathrm{u}(1-\alpha/2)\sqrt{\mathrm{var}(\hat{\beta}_j)}\right), \exp\left(\hat{\beta}_j + \mathrm{u}(1-\alpha/2)\sqrt{\mathrm{var}(\hat{\beta}_j)}\right) \right\rangle.$$

Můžeme si všimnout, že kvůli exponenciální transformaci nejsou intervalové odhady poměrů šancí symetrické kolem jejich bodových odhadů.

1.4. Testování významnosti regresních parametrů

V předešlé podkapitole jsme si ukázali, že odhady regresních parametrů získáme metodou maximální věrohodnosti. Nyní nás bude přirozeně zajímat, zda jsou tyto parametry signifikantně nenulové. Jinými slovy, zda vůbec mají vysvětlující proměnné významný vliv na vysvětlovanou proměnnou. Hypotézy o parametrech, jejichž odhady získáváme pomocí metody maximální věrohodnosti, většinou testujeme pomocí testu poměru věrohodností, Waldova testu nebo skórového testu. My si zde představíme první dva zmíněné.

Nejprve budeme testovat celkový vliv vysvětlujících proměnných na vysvětlovanou proměnnou, tj. budeme testovat hypotézu

$$\mathrm{H}_0 \colon \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

$$\mathrm{H}_A\colon \beta_j \neq 0 \text{ pro alespoň jedno } j.$$

Takto formulovanou hypotézu můžeme testovat pomocí testu poměru věrohodností (*likelihood ratio test*). Testová statistika je ve tvaru

$$G = -2\ln\left[\frac{\text{věrohodnost nulového modelu}}{\text{věrohodnost úplného modelu}}\right], \tag{1.8}$$

kde nulový model je model bez vysvětlujících proměnných, tj. pouze s absolutním členem β_0, a úplný model je model se všemi vysvětlujícími proměnnými.

Označme si n_1 jako počet pozorování, kde nastal zkoumaný jev, tj. $n_1 = \sum y_i$, a n_0 jako počet pozorování, kdy zkoumaný jev nenastal, tj. $n_0 = \sum(1 - y_i)$. Pro získání odhadu parametru β_0 metodou maximální věrohodnosti vyřešíme rovnici

$$\begin{aligned}
\sum_{i=1}^{n}[y_i - \pi(\boldsymbol{x}_i)] &= 0\\
\sum_{i=1}^{n} y_i - n\frac{\exp(\beta_0)}{1+\exp(\beta_0)} &= 0\\
n\frac{\exp(\beta_0)}{1+\exp(\beta_0)} &= n_1\\
n\exp(\beta_0) &= n_1 + n_1\exp(\beta_0)\\
n_0\exp(\beta_0) &= n_1\\
\hat{\beta}_0 &= \ln\left(\frac{n_1}{n_0}\right).
\end{aligned}$$

Odsud pak dosazením do (1.4) odvodíme odhad $\hat{\pi}$

$$\hat{\pi} = \frac{\exp(\hat{\beta}_0)}{1+\exp(\hat{\beta}_0)} = \frac{n_1/n_0}{1+n_1/n_0} = \frac{n_1/n_0}{n/n_0} = \frac{n_1}{n},$$

který již nezávisí na hodnotách $\boldsymbol{x}_i$. Po dosazení do testové statistiky (1.8) získáme

$$G = -2\ln\left[\frac{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^{n}\hat{\pi}(\boldsymbol{x}_i)^{y_i}[1-\hat{\pi}(\boldsymbol{x}_i)]^{(1-y_i)}}\right],$$

po úpravě

$$G = 2\left\{\sum_{i=1}^{n}\left[y_i \ln[\hat{\pi}(\boldsymbol{x}_i)] + (1-y_i)\ln[1-\hat{\pi}(\boldsymbol{x}_i)]\right] - \left[n_1 \ln(n_1) + n_0 \ln(n_0) - n\ln(n)\right]\right\}.$$

Pro dostatečně velký počet pozorování má testová statistika G za platnosti nulové hypotézy χ^2 rozdělení s k stupni volnosti. Pokud se tedy testová statistika realizuje hodnotou $g > \chi^2_k(1-\alpha)$, kde $\chi^2_k(1-\alpha)$ značí $(1-\alpha)$-kvantil χ^2 rozdělení s k stupni volnosti, zamítneme nulovou hypotézu na hladině významnosti α.

Výše uvedený test je speciálním případem tzv. testu podmodelu. Mějme model M s m regresními parametry (v našem případě $m = k+1$). Naším cílem je testovat nulovou hypotézu, že l-tice regresních parametrů modelu M, $l < m$, je nulová. Za platnosti nulové hypotézy tedy dostaneme zjednodušený model M^* s $m-l$ regresními parametry, který označujeme jako podmodel modelu M. Testová statistika

$$G = -2\ln\left[\frac{\text{věrohodnost modelu } M^*}{\text{věrohodnost modelu } M}\right],$$

má pro dostatečně velký počet pozorování za platnosti nulové hypotézy přibližně χ^2 rozdělení s l stupni volnosti.

Pro testování významnosti jednotlivých regresních parametrů se často místo testu poměru věrohodností používá i tzv. Waldův test. Nulovou a alternativní hypotézu testu významnosti parametru β_j pro dané $j \in \{0, 1, \ldots, k\}$ budeme formulovat ve tvaru

$$\mathrm{H}_0 \colon \beta_j = 0$$
$$\mathrm{H}_A \colon \beta_j \neq 0.$$

Waldův test je založen na Waldově testové statistice

$$W_j = \frac{\hat{\beta}_j}{\sqrt{\mathrm{var}(\hat{\beta}_j)}},$$

která má pro dostatečně velký počet pozorování za platnosti nulové hypotézy přibližně normované normální rozdělení. Pokud pro realizaci w_j testové statistiky W_j bude platit, že $|w_j| \geq \mathrm{u}(1-\alpha/2)$, zamítneme nulovou hypotézu na hladině významnosti α a řekneme, že regresní parametr β_j je signifikantně nenulový.

Pro odhadování regresních parametrů, testování významnosti a jiné nástroje logistické regrese se v programu R nejčastěji využívá funkce `glm`, kde je potřeba nastavit parametr `family` na hodnotu `binomial`.

1.5. Výběr proměnných a hodnocení modelu

V předešlé podkapitole jsme si ukázali, jak testovat významnost regresních parametrů. Naším cílem je získat kvalitní model, ve kterém jsou všechny proměnné, které mají vliv na vysvětlovanou proměnnou. Na druhou stranu není žádoucí do modelu zbytečně zahrnovat nesignifikantní proměnné. Přístupů, jak vybrat co nejvhodnější složení proměnných do modelu, je mnoho. Jedním z nich je tzv. kroková selekce (*stepwise selection*).

V rámci krokové selekce postupně odebíráme, či přidáváme proměnné do modelu tak, abychom optimalizovali hodnotu informačního kritéria. Nejpoužívanějším informačním kritériem je Akaikeho informační kritérium (AIC - *Akaike information criterion*) definované jako

$$AIC = 2k - 2\ln[L(\hat{\boldsymbol{\beta}})] = 2k - 2l(\hat{\boldsymbol{\beta}}).$$

Po dosazení výrazu (1.6) dostaneme

$$AIC = 2k - 2\sum_{i=1}^{n}\Big\{y_i \ln[\hat{\pi}(\boldsymbol{x}_i)] + (1-y_i)\ln[1-\hat{\pi}(\boldsymbol{x}_i)]\Big\}.$$

Proces sestupné krokové selekce (*backward stepwise selection*) začínáme s úplným modelem, tedy s modelem, ve kterém jsou zahrnuty všechny vysvětlující proměnné, které máme k dispozici. Následně z modelu vyřadíme vždy takovou vysvětlující proměnnou, jejíž odebrání z modelu zapříčiní největší pokles hodnoty

AIC. Proces ukončíme, pokud odebrání jakékoliv vysvětlující proměnné nevede k poklesu hodnoty AIC.

Opačným přístupem je vzestupná kroková selekce (*forward stepwise selection*). Proces začínáme s nulovým modelem, tedy s modelem, který obsahuje pouze absolutní člen. Následně do modelu přidáme vždy takovou vysvětlující proměnnou, jejíž přidání do modelu vede k největšímu poklesu hodnoty AIC. Proces ukončíme, pokud přidání jakékoliv další vysvětlující proměnné nevede k poklesu hodnoty AIC.

Kombinací obou dvou předešlých přístupů je obousměrná kroková selekce (*bi-directional stepwise selection*). Začínáme stejně jako u vzestupné selekce s nulovým modelem. V každém kroku pak buď vyřadíme, nebo přidáme do modelu vysvětlující proměnnou tak, aby hodnota AIC co nejvíce poklesla. Proces opět ukončíme, pokud již není možné hodnotu AIC snížit.

V programu R lze krokovou selekci zrealizovat pomocí funkce `step`, ve které nastavíme parametr `direction` na `backward`, `forward` nebo `both`.

Jakmile máme sestavený model s optimálním složením proměnných, je načase provést diagnostiku a hodnocení tohoto modelu. Jako první zkontrolujeme splnění předpokladů modelu. Prvním z předpokladů je linearita vztahu mezi kvantitativními vysvětlujícími proměnnými a logitem pravděpodobnosti nastání jevu daného vysvětlovanou proměnnou, $g(\boldsymbol{x})$. Po dosazení odhadnutých regresních parametrů do vztahu (1.3) získáme pro n pozorování vyrovnané hodnoty $\hat{g}(\boldsymbol{x}_i)$, $i = 1, \ldots, n$. Předpoklad linearity pak dokážeme ověřit vizuálně pomocí grafu závislosti těchto vyrovnaných hodnot na napozorovaných hodnotách kvantitativních proměnných.

Dalším předpokladem je lineární nezávislost vysvětlujících proměnných. Porušení tohoto předpokladu nazýváme problémem multikolinearity. Závažnost multikolinearity můžeme měřit např. pomocí faktoru zvětšení rozptylu (VIF - *Variance Influence Factor*) definovaného jako

$$VIF_j = \frac{1}{1 - R_j^2}, \quad j = 1, \ldots, k,$$

kde R_j^2 představuje index determinace pomocného modelu, ve kterém je j-tá proměnná vysvětlována pomocí zbylých $k-1$ vysvětlujících proměnných. V případě lineární nezávislosti proměnných je VIF roven jedné. Obecně uznávaným pravidlem je, že hodnoty VIF větší než 5 poukazují na slabou multikolinearitu a hodnoty VIF větší než 10 poukazují na silnou multikolinearitu.

Výsledky modelu mohou být ovlivněny i tzv. vlivnými pozorováními. Ta mohou být detekována např. pomocí Cookovy vzdálenosti a standardizovaných reziduí. Pro více informací odkazujeme čtenáře na zdroj [7].

Pro posouzení celkové správnosti modelu můžeme použít tzv. Hosmerův-Lemeshowův test dobré shody. Dosazením vyrovnaných hodnot $\hat{g}(\boldsymbol{x}_i)$ do vzorce (1.4) dostaneme odhady pravděpodobností $\hat{\pi}(\boldsymbol{x}_i)$. Tyto odhady uspořádáme vzestupně a následně rozdělíme podle decilů do 10 skupin o velikosti $n'_h = n/10$, kde $h = 1, \ldots, 10$ [1]. V první skupině bude tedy 10 % nejmenších odhadů $\hat{\pi}(\boldsymbol{x}_i)$ a v desáté skupině 10 % největších odhadů $\hat{\pi}(\boldsymbol{x}_i)$. V každé skupině následně spočítáme aritmetický průměr hodnot $\hat{\pi}(\boldsymbol{x}_i)$, který označíme jako $\overline{\pi}_h$ a který představuje očekávanou proporci nastání zkoumaného jevu v h-té skupině. Nulovou hypotézou je shoda očekávaných a napozorovaných proporcí nastání zkoumaného jevu ve všech skupinách, což odpovídá správně specifikovanému modelu. Alternativní hypotézou je pak tvrzení, že alespoň v jedné ze skupin se očekávaná proporce liší od té napozorované. V rámci výpočtu testové statistiky budeme místo proporcí pracovat s četnostmi. Označme proto napozorovanou četnost nastání zkoumaného jevu v h-té skupině jako o_h. Očekávanou četnost nastání zkoumaného jevu v h-té skupině označme jako e_h a získáme ji jako $e_h = n'_h \overline{\pi}_h$. Testová statistika pro Hosmerův-Lemeshowův test je ve tvaru

$$\hat{C} = \sum_{h=1}^{10} \frac{(o_h - e_h)^2}{e_h(1 - \overline{\pi}_h)}$$

a její přibližné rozdělení je χ^2 rozdělení s 8 stupni volnosti. Ve prospěch alternativní hypotézy svědčí velké hodnoty testové statistiky. Pokud se tedy tes-

[1] V případě neceločíselného výsledku nejprve spočítáme celou dolní část, tedy $d = \lfloor n/10 \rfloor$. Pak $n_h = d + 1$ pro $h = 1, \ldots, (n - 10d)$ a pro zbylé skupiny je $n_h = d$.

tová statistika realizuje hodnotou $\hat{c} > \chi^2_8(1-\alpha)$, zamítneme nulovou hypotézu o správnosti modelu na hladině významnosti α.

Kapitola 2

Vícenásobná logistická regrese se smíšenými efekty

V minulé kapitole jsme si představili vícenásobný logistický model, tedy speciální případ zobecněných regresních modelů. Předpokladem klasického logistického modelu je statistická nezávislost všech pozorování. Pokud tedy máme v datovém souboru strukturu, nemůžeme klasický logistický model použít. Zmíněnou strukturou mohou být například nezávislé shluky pozorování, uvnitř kterých jsou ale pozorování závislá. Příkladem mohou být data o pacientech, u kterých zaznamenáme opakovaná měření. Jednotliví pacienti jsou mezi sebou nezávislí, opakovaná měření u jednoho pacienta už ale závislá jsou. Zanedbání takové struktury v datech a použití modelu s porušenými předpoklady může vést k naprosto chybným závěrům.

Jednou možností, jak se se strukturou v datech vypořádat, je agregace dat například pomocí aritmetického průměru či mediánu. Pro každého pacienta bychom tak měli jedno pozorování nezávislé na pozorováních ostatních pacientů. Nevýhodou tohoto přístupu je ale ztráta cenné informace. Proto se k analýze datových souborů se strukturou stále častěji využívají zobecněné regresní modely se smíšenými efekty (GLMM - *Generalized linear mixed models*). Ty vzniknou přidáním náhodných efektů k již stávajícím pevným efektům v modelu. My se konkrétně zaměříme na vícenásobný logistický model s náhodným absolutním členem, speciální případ zobecněných smíšených regresních modelů.

Nyní si nejprve vysvětlíme rozdíl mezi pevným a náhodným efektem, následně zavedeme logistický model s náhodným absolutním členem a nakonec uvedeme metody pro odhad a testování parametrů modelu. Hlavními zdroji pro tuto kapitolu jsou [3], [4], [9], [10], [13], [14], [16], [19] a [20].

2.1. Pevné a náhodné efekty

Jak už jsme naznačili v úvodu této kapitoly, smíšené regresní modely nesou své jméno podle kombinace pevných a náhodných efektů v modelu. Pevným efektem (*fixed effect*) označujeme nenáhodnou vysvětlující proměnnou, na kterou jsme zvyklí z klasických regresních modelů. Může se jednat o kvantitativní proměnnou, jako je např. věk pacienta, nebo se může jednat o kvalitativní proměnnou (faktor), jako je např. pohlaví. Náhodným efektem (*random effect*) je kvalitativní vysvětlující proměnná, jejíž úrovně jsou náhodným výběrem z široké populace takových úrovní. Typickým příkladem náhodného efektu je např. označení pacienta, v datech tedy máme pozorování náhodně vybrané skupiny ze všech pacientů trpících určitou nemocí. My si nyní popíšeme rozdíly mezi pevnými faktory a náhodnými efekty.

Pevný faktor sestává většinou z malého množství kategorií, které do modelu zahrnujeme pomocí umělých proměnných (viz podkapitola 1.2). Náhodný efekt má často více než pět úrovní, které tvoří v datech dříve zmiňovanou strukturu. Pokud máme například opakovaná měření u 150 pacientů, náhodným efektem bude označení pacienta se 150 úrovněmi. Zahrnutí takové kategoriální proměnné do modelu pomocí umělých proměnných by vyžadovalo velké množství parametrů, což je nežádoucí.

Dalším rozdílem je, že u pevného faktoru jsou kategorie jasně určené a jejich význam je stálý. Pokud je tedy statistický subjekt v jednom datovém souboru označen jako *muž*, bude takto označen i v jiném datovém souboru se stejným významem. Naproti tomu označení úrovně náhodného efektu je většinou pouze identifikační číslo, které nemá samo o sobě žádný význam. Subjekt, který je označený v jednom souboru jako *pacient 1*, může být v jiném datovém souboru

označen např. jako *pacient 2*.

Navíc u pevného faktoru cíleně odlišujeme jednotlivé kategorie. Jinými slovy, zajímá nás např. jak velký je rozdíl mezi efekty pro muže a ženy. Rozlišení úrovní náhodného efektu naproti tomu není středem našeho zájmu. Pokud máme v datech pozorování o 150 pacientech, není naším cílem zjistit, jaký je rozdíl mezi jednotlivými pacienty. Do modelu většinou náhodný efekt zahrnujeme pouze pro podchycení variability.

Na základě výše uvedených informací můžeme shrnout, že vhodnými daty pro smíšené regresní modely jsou longitudinální data, panelová data, data s opakovanými měřeními a obecně data s nezávislými skupinami závislých pozorování.

Rozhodování, zda je daná proměnná pevným či náhodným efektem, nemusí být vždy jednoduché. Zásadní je především znalost prostředí a okolností, ze kterých data pochází. Proměnná, která je v jedné studii v pozici pevného efektu, může být v jiné studii v pozici efektu náhodného. Ještě jednou ale musíme zdůraznit, že zanedbání přítomnosti náhodného efektu může zásadním způsobem ovlivnit výsledky analýzy.

2.2. Vícenásobný logistický model s náhodným absolutním členem

Vícenásobný logistický model s náhodným absolutním členem je jedním z nejjednodušších, ale zároveň velmi často používaných zobecněných lineárních modelů se smíšenými efekty. Při zavedení modelu budeme vycházet z klasického logistického modelu (1.3). Mějme tedy náhodnou závisle proměnnou Y s alternativním rozdělením a k pevných efektů $x_1, x_2, \ldots, x_k$.

Smíšený logistický model s náhodným absolutním členem pro datový soubor se strukturou N nezávislých skupin pozorování zapíšeme ve tvaru

$$g(\boldsymbol{x}_{ij}) = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \cdots + \beta_k x_{ijk} + u_j, \quad i = 1, 2, \ldots, n_j, \quad j = 1, 2, \ldots, N,$$

kde $\boldsymbol{x}_{ij}$ značí i-té pozorování pevných efektů v j-té skupině. Dále pak n_j značí počet pozorování v j-té skupině, neomezujeme se tedy pouze na skupiny se

stejným počtem pozorování. Pokud navážeme na příklad datového souboru opakovaných měření pacientů z předchozí podkapitoly, $N = 150$ a $\boldsymbol{x}_{ij}$ je i-té pozorování u j-tého pacienta.

Náhodný efekt do modelu zahrnujeme v podobě náhodného členu u_j. O něm předpokládáme, že se řídí normálním rozdělením s nulovou střední hodnotou a rozptylem σ^2, tj. $u_j \sim N(0, \sigma^2)$, a že pro všechna $j = 1, \ldots, N$ jsou efekty u_j nezávislé. Pro j-tou skupinu pozorování je náhodný absolutní člen ve tvaru $\beta_0 + u_j$. Parametr β_0 je v kontextu smíšených modelů označován jako absolutní člen společný pro celou populaci (*population-averaged intercept*).

Jako spojovací funkci opět uvažujeme logitovou funkci

$$g(\boldsymbol{x}_{ij}) = \ln\left(\frac{\pi(\boldsymbol{x}_{ij})}{1 - \pi(\boldsymbol{x}_{ij})}\right), \tag{2.1}$$

kde $\pi(\boldsymbol{x}_{ij})$ je podmíněná střední hodnota náhodné veličiny Y,

$$\pi(\boldsymbol{x}_{ij}) = \mathrm{E}(Y_{ij}|\boldsymbol{x}_{ij}, u_j) = P(Y_{ij} = 1|\boldsymbol{x}_{ij}, u_j).$$

Předpokládáme, že pozorování Y_{ij} se řídí podmíněným alternativním rozdělením, tj. $Y_{ij}|u_j \sim Alt(p_{ij})$. Navíc předpokládáme, že podmíněně vzhledem k náhodné složce u_j jsou pozorování Y_{ij} nezávislá v rámci jedné skupiny i mezi jednotlivými skupinami, tedy pro všechna $i = 1, \ldots, n_j$ a $j = 1, \ldots, N$.

Interpretace regresních parametrů zůstává téměř stejná jako u klasického logistického modelu, tj. pomocí poměru šancí. Je nyní ale zapotřebí dodat, že dané hodnoty jsou podmíněné hodnotou realizace náhodné složky u_j.

2.3. Odhady parametrů logistického modelu s náhodným absolutním členem

Nyní se budeme zabývat metodami pro odhady parametrů smíšeného logistického modelu s náhodným absolutním členem. Nejprve se zaměříme na odhady pevných efektů a parametru σ^2, poté si ukážeme postup pro predikci náhodných efektů.

2.3.1. Odhady pevných efektů a parametru σ^2

Jako u klasického logistického modelu, regresní parametry $\beta_0, \beta_1, \ldots, \beta_k$ a σ^2 budeme odhadovat pomocí metody maximální věrohodnosti. Nejprve si sestavíme věrohodnostní funkci

$$
\begin{aligned}
L(\boldsymbol{\beta}, \sigma^2) &= P(Y = y|\boldsymbol{\beta}, \sigma^2) \\
&= \prod_{j=1}^{N} \int_{-\infty}^{\infty} \prod_{i=1}^{n_j} f(y_{ij}|\boldsymbol{\beta}, u_j) f(u_j|\sigma^2)\, du_j \\
&= \prod_{j=1}^{N} \int_{-\infty}^{\infty} \prod_{i=1}^{n_j} \pi(\boldsymbol{x}_{ij})^{y_{ij}} [1 - \pi(\boldsymbol{x}_{ij})]^{(1-y_{ij})} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{u_j^2}{2\sigma^2}\right\} du_j \\
&= \prod_{j=1}^{N} \int_{-\infty}^{\infty} \prod_{i=1}^{n_j} \left[\frac{\exp\left[g(\boldsymbol{x}_{ij})\right]}{1 + \exp\left[g(\boldsymbol{x}_{ij})\right]}\right]^{y_{ij}} \left[\frac{1}{1 + \exp\left[g(\boldsymbol{x}_{ij})\right]}\right]^{(1-y_{ij})} \\
&\qquad \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{u_j^2}{2\sigma^2}\right\} du_j \\
&= \prod_{j=1}^{N} \int_{-\infty}^{\infty} \prod_{i=1}^{n_j} \frac{\exp\left[y_{ij} g(\boldsymbol{x}_{ij})\right]}{1 + \exp\left[g(\boldsymbol{x}_{ij})\right]} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{u_j^2}{2\sigma^2}\right\} du_j \\
&= (2\pi\sigma^2)^{-\frac{N}{2}} \prod_{j=1}^{N} \prod_{i=1}^{n_j} \exp\left[y_{ij}\beta_0 + y_{ij} \sum_{s=1}^{k} \beta_s x_{ijs}\right] \int_{-\infty}^{\infty} \frac{\exp\left[y_{ij} u_j - \frac{u_j^2}{2\sigma^2}\right]}{1 + \exp\left[g(\boldsymbol{x}_{ij})\right]} du_j,
\end{aligned}
$$

kde $f(y_{ij}|\boldsymbol{\beta}, u_j) = P(Y_{ij} = y_{ij}|\boldsymbol{\beta}, u_j)$ a $f(u_j|\sigma^2)$ je hustota rozdělení pravděpodobnosti náhodného efektu, tj. hustota normálního rozdělení. Zintegrováním výrazu jsme odbourali náhodný efekt u_j a získali tak marginální rozdělení náhodné veličiny Y_{ij}. Logaritmickou věrohodnostní funkci pak získáme jako logaritmus funkce $L(\boldsymbol{\beta}, \sigma^2)$

$$
\begin{aligned}
l(\boldsymbol{\beta}, \sigma^2) &= \ln[L(\boldsymbol{\beta}, \sigma^2)] \\
&= -\frac{N}{2} \ln(2\pi\sigma^2) + \sum_{i=1}^{n_j} \sum_{j=1}^{N} \left[y_{ij}\beta_0 + y_{ij} \sum_{s=1}^{k} \beta_s x_{ijs}\right] \\
&\quad + \sum_{j=1}^{N} \ln \int_{-\infty}^{\infty} \exp\left\{\sum_{i=1}^{n_j} y_{ij} u_j - \frac{u_j^2}{2\sigma^2} - \sum_{i=1}^{n_j} \ln\left[1 + \exp\left[g(\boldsymbol{x}_{ij})\right]\right]\right\} du_j.
\end{aligned}
$$

Označme pro jednoduchost

$$h_j(\boldsymbol{\beta}, u_j) = \sum_{i=1}^{n_j} y_{ij} u_j - \frac{u_j^2}{2\sigma^2} - \sum_{i=1}^{n_j} \ln\left[1 + \exp\left[g(\boldsymbol{x}_{ij})\right]\right],$$

potom

$$l(\boldsymbol{\beta}, \sigma^2) = -\frac{N}{2}\ln(2\pi\sigma^2) + \sum_{i=1}^{n_j}\sum_{j=1}^{N}\left[y_{ij}\beta_0 + y_{ij}\sum_{s=1}^{k}\beta_s x_{ijs}\right]$$
$$+ \sum_{j=1}^{N} \ln \int_{-\infty}^{\infty} \exp\left[h_j(\boldsymbol{\beta}, u_j)\right]\, du_j.$$

Nyní zderivujeme logaritmickou věrohodnostní funkci podle parametrů $\beta_0, \ldots, \beta_k$ a σ^2.

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \beta_0} = \sum_{i=1}^{n_j}\sum_{j=1}^{N} y_{ij} - \sum_{j=1}^{N} \frac{\mathrm{I}_{j1}}{\mathrm{I}_{j2}},$$

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \beta_s} = \sum_{i=1}^{n_j}\sum_{j=1}^{N} y_{ij} x_{ijs} - \sum_{j=1}^{N} \frac{\mathrm{I}_{js3}}{\mathrm{I}_{j2}}, \quad s = 1, \ldots, k,$$

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{j=1}^{N} \frac{\mathrm{I}_{j4}}{\mathrm{I}_{j2}},$$

kde

$$\mathrm{I}_{j1} = \int_{-\infty}^{\infty} \sum_{i=1}^{n_j} \exp\left[h_j(\boldsymbol{\beta}, u_j)\right] \frac{\exp\left[g(\boldsymbol{x}_{ij})\right]}{1 + \exp\left[g(\boldsymbol{x}_{ij})\right]}\, du_j,$$

$$\mathrm{I}_{j2} = \int_{-\infty}^{\infty} \exp\left[h_j(\boldsymbol{\beta}, u_j)\right]\, du_j,$$

$$\mathrm{I}_{js3} = \int_{-\infty}^{\infty} \sum_{i=1}^{n_j} \exp\left[h_j(\boldsymbol{\beta}, u_j)\right] \frac{\exp\left[g(\boldsymbol{x}_{ij})\right]}{1 + \exp\left[g(\boldsymbol{x}_{ij})\right]} x_{ijs}\, du_j,$$

$$\mathrm{I}_{j4} = \int_{-\infty}^{\infty} \exp\left[h_j(\boldsymbol{\beta}, u_j)\right] u_j^2\, du_j.$$

Následně položíme výrazy rovny nule a vyřešíme soustavu rovnic. Problém maximalizace je na první pohled značně komplikovanější než u klasického logistického modelu, a to především díky integrálům, které musí být evaluovány

numericky. Jednou z metod, které se k jejich výpočtu využívají, je Gaussova-Hermitova kvadratura (GHQ - *Gauss-Hermite Quadrature*).

Gaussova-Hermitova kvadratura

Integrál lze interpretovat jako nekonečný vážený součet. Kvadratura je založená na aproximaci integrálu pomocí konečného váženého součtu hodnot integrované funkce, vyčíslené na množině hodnot proměnné, přes kterou integrujeme.

Mějme integrál

$$\int_{-\infty}^{\infty} \exp(-x^2) f(x)\, dx,$$

kde $f(x)$ je hladká funkce, která lze dostatečně dobře aproximovat polynomem. Gaussova-Hermitova kvadratura aproximuje daný integrál jako

$$\int_{-\infty}^{\infty} \exp(-x^2) f(x)\, dx \approx \sum_{r=1}^{R} p_r^* f(a_r^*),$$

kde p_r^* nazýváme váhy a body a_r^* nazýváme uzly. Jednotlivé volby vah a uzlů nazýváme kvadraturní pravidla a dají se vyčíst z tabulek, nebo je obdržet softwarem. Metoda GHQ je pro účely odhadu regresních parametrů zobecněného smíšeného modelu výhodná, jelikož váhová funkce $\exp(-x^2)$ je proporcionální k hustotě normálního rozdělení, jejíž funkční předpis se objevuje v integrálech $\mathrm{I}_{j1}, \mathrm{I}_{j2}$, I_{js3} a I_{j4}.

S rostoucím počtem uzlů R se přesnost aproximace mírně zlepšuje, pro účely zobecněných smíšených modelů se doporučuje volit $R \geq 20$. Nutno ale podotknout, že v některých případech ani velké hodnoty R nezaručují dostatečně přesnou aproximaci integrálů. Problémy s přesností se navíc zhoršují s rostoucí velikostí skupin pozorování a s rostoucím rozptylem náhodné složky u_j.

Alternativní metodou pro aproximaci integrálů je Adaptivní Gaussova kvadratura (AGQ - *Adaptive Gauss Quadrature*). Adaptivní Gaussova kvadratura je odvozena od metody GHQ, rozdíl ale spočívá ve volbě uzlů. U metody GHQ jsou uzly pevně volené nezávisle na tvaru integrované funkce. U metody AGQ jsou uzly posunuty a přeškálovány tak, aby ležely pod vrcholem integrované funkce.

Díky tomu je metoda mnohem přesnější než GHQ i pro mnohem menší počet uzlů. Na druhé straně je ale metoda AGQ výpočetně náročnější, tím pádem i pomalejší, celkově je ale doporučováno využít spíše metodu AGQ.

Alternativní metodou pro numerické integrování je např. metoda Monte Carlo spolu s EM algoritmem (*Expectation-Maximization algorithm*). Jak už jsme ale naznačili, nevýhodou numerického integrování je velká výpočetní náročnost. Pro složitější modely, jako jsou např. modely s větším počtem náhodných efektů, kde se objevují vícedimenzionální integrály, je proto žádoucí použití jiného přístupu. Nejpoužívanějším z těchto přístupů je odhadování založené na aproximaci věrohodnostní funkce pomocí LaPlaceovy aproximace, speciálně pak z ní odvozená metoda penalizované kvazi-věrohodnosti (PQL - *Penalized Quasi-Likelihood*). Pro další podrobnosti o všech výše zmíněných metodách odkazujeme čtenáře na zdroj [3].

Označme pro tuto chvíli odhady vektorového parametru $\boldsymbol{\theta} = (\beta_0, \ldots, \beta_k, \sigma^2)'$ jako $\hat{\boldsymbol{\theta}} = (\hat{\beta}_0, \ldots, \hat{\beta}_k, \hat{\sigma}^2)'$. Obdobně jako u logistické regrese varianční matici vektoru $\hat{\boldsymbol{\theta}}$ získáme jako inverzi Fisherovy informační matice $\boldsymbol{I}(\boldsymbol{\theta})$ evaluovanou v bodě $\hat{\boldsymbol{\theta}}$, tj. $\mathrm{var}(\hat{\boldsymbol{\theta}}) = \boldsymbol{I}^{-1}(\hat{\boldsymbol{\theta}})$. Prvky matice $\boldsymbol{I}(\boldsymbol{\theta})$ jsou opět dány jako

$$\boldsymbol{I}(\boldsymbol{\theta}) = (i_{js})_{j,s=1}^{k+2}, \qquad i_{js} = -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\theta_j \partial\theta_s}.$$

Tvary druhých parciálních derivací funkce $l(\boldsymbol{\theta})$ zde odvozovat nebudeme, čtenář je nalezne např. ve zdroji [3].

2.3.2. Predikce náhodných efektů

Nyní se zaměříme na predikce náhodných efektů. Jak již bylo zmíněno v podkapitole 2.1, náhodné efekty většinou nejsou středem zájmu. I přesto, že jejich hodnoty nebudeme dále dopodrobna zkoumat a porovnávat, jejich vyčíslení je důležité pro další účely, jako např. pro výpočet vyrovnaných hodnot. Nejčastěji se pro predikci náhodných efektů využívá metoda sdružené maximální věrohodnosti (JML - *Joint Maximum Likelihood*), která je založená na společném odhadování

pevných a náhodných efektů. Zaveďme nyní sdruženou věrohodnostní funkci

$$L_{JML}(\boldsymbol{\beta}, u_j) = \prod_{j=1}^{N}\prod_{i=1}^{n_j} f(y_{ij}, u_j|\boldsymbol{\beta}, \sigma^2) = \prod_{j=1}^{N}\prod_{i=1}^{n_j} f(y_{ij}|\boldsymbol{\beta}, \sigma^2) f(u_j|y_{ij}, \boldsymbol{\beta}, \sigma^2)$$

$$= \prod_{j=1}^{N}\prod_{i=1}^{n_j} \exp\left\{y_{ij}\ln[\pi(\boldsymbol{x}_{ij})] + (1-y_{ij})\ln[1-\pi(\boldsymbol{x}_{ij})]\right\}$$

$$\times (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{u_j^2}{2\sigma^2}\right\}.$$

Zlogaritmováním získáme sdruženou logaritmickou věrohodnostní funkci

$$l_{JML}(\boldsymbol{\beta}, u_j) = -\frac{N}{2}\ln(2\pi\sigma^2)$$

$$+ \sum_{j=1}^{N}\sum_{i=1}^{n_j}\left\{y_{ij}\ln[\pi(\boldsymbol{x}_{ij})] + (1-y_{ij})\ln[1-\pi(\boldsymbol{x}_{ij})] - \frac{u_j^2}{2\sigma^2}\right\}.$$

Odhady $\hat{\boldsymbol{\beta}}$ a $\hat{u}_j$ nalezneme vyřešením soustavy rovnic

$$\frac{\partial l_{JML}(\boldsymbol{\beta}, u_j)}{\partial \beta_s} = 0, \qquad s = 0, 1, \ldots, k,$$

$$\frac{\partial l_{JML}(\boldsymbol{\beta}, u_j)}{\partial u_j} = 0, \qquad j = 1, \ldots, N.$$

V praxi počet náhodných efektů často dosahuje řádu desítek až stovek, musíme tedy současně řešit velké množství nelineárních rovnic. Proto Jiang [9] k řešení nedoporučuje v praxi často využívaný Newtonův-Raphsonův algoritmus, který je pro velký počet rovnic neefektivní a velmi pomalý. Navíc je také velmi citlivý na počáteční hodnoty, jejichž volba může být pro takové množství efektů obtížná. Místo Newtonova-Raphsonova algoritmu navrhuje použití Gaussova-Seidlova algoritmu.

Nakonec ještě poznamenáme, že odhady pevných efektů, které metodou sdružené maximální věrohodnosti získáme spolu s predikcemi náhodných efektů, nejsou konzistentní. Navíc není zaručené, že budou asymptoticky nejlepšími odhady parametru $\boldsymbol{\beta}$. Proto se pro odhadování pevných efektů využívají metody zmíněné výše v podkapitole 2.3.1.

2.4. Testování významnosti regresních parametrů logistického modelu s náhodným absolutním členem

Stejně jako u klasického logistického modelu, i u smíšeného logistického modelu s náhodným absolutním členem se nejčastěji k testování hypotéz využívá test poměru věrohodností a Waldův test.

Pro test celkového vlivu vysvětlujících proměnných, tj. pro test hypotézy

$$\mathrm{H}_0\colon \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

$$\mathrm{H}_A\colon \beta_s \neq 0 \text{ pro alespoň jedno } s \in \{1, \ldots, k\},$$

využijeme opět testovou statistiku testu poměru věrohodností

$$G = -2 \ln \left[\frac{\text{věrohodnost nulového modelu}}{\text{věrohodnost úplného modelu}} \right],$$

která má pro dostatečně velký počet pozorování za platnosti nulové hypotézy přibližně χ^2 rozdělení s k stupni volnosti.

Obdobně i pro test podmodelu M^* modelu M, tj. pro test hypotézy, že l-tice regresních parametrů modelu M je nulová, použijeme testovou statistiku

$$G = -2 \ln \left[\frac{\text{věrohodnost modelu } M^*}{\text{věrohodnost modelu } M} \right]. \tag{2.2}$$

Ta má pro dostatečně velký počet pozorování za platnosti nulové hypotézy opět přibližně χ^2 rozdělení s l stupni volnosti.

Pro testování významnosti jednotlivých regresních parametrů se kromě testu poměru věrohodností používá i Waldův test. Mějme tedy pro dané $s \in \{0, 1, \ldots, k\}$ hypotézu ve tvaru

$$\mathrm{H}_0\colon \beta_s = 0$$

$$\mathrm{H}_A\colon \beta_s \neq 0.$$

Waldova testová statistika

$$W_s = \frac{\hat{\beta}_s}{\sqrt{\mathrm{var}(\hat{\beta}_s)}}$$

má pro dostatečně velký počet pozorování za platnosti nulové hypotézy přibližně normované normální rozdělení. Waldův test je často upřednostňován pro testování významnosti jednotlivých regresních parametrů před testem poměru věrohodností, a to kvůli menší výpočetní náročnosti. Musíme si totiž uvědomit, že na rozdíl od Waldovy statistiky, pro výpočet hodnoty testové statistiky G musíme sestavit i zjednodušený model, což je kvůli numerické evaluaci integrálů velmi výpočetně náročné.

Nyní se budeme zabývat signifikancí náhodného efektu. Jestliže máme náhodný efekt u_j, pro který platí, že $u_j \sim N(0, \sigma^2)$, test signifikance takového efektu odpovídá testu nulovosti parametru σ^2. Mějme tedy hypotézu ve tvaru

$$\mathrm{H}_0 \colon \sigma^2 = 0$$
$$\mathrm{H}_A \colon \sigma^2 > 0.$$

Pro testování můžeme opět použít test poměru věrohodností a analogii k testové statistice (2.2). Vidíme, že díky nezápornosti rozptylu se jedná o jednostrannou hypotézu. Právě proto, že se za platnosti nulové hypotézy pohybujeme na samé hranici přípustných hodnot, kterých obecně rozptyl může nabývat, mnoho zdrojů nabádá k opatrnosti při interpretaci výsledku testu. Podle zdroje [19] mohou být p-hodnoty tohoto testu poměru věrohodností nadhodnocené, což může vést k nezamítnutí neplatné nulové hypotézy (chyba 2. druhu). Proto se někdy doporučuje porovnávat s hladinou významnosti polovinu výsledné p-hodnoty.

Nutno navíc podotknout, že tento test lze použít pouze u modelů s vícero náhodnými efekty. V našem případě, kdy máme pouze jeden náhodný efekt (náhodný absolutní člen), získáme za platnosti nulové hypotézy klasický logistický model. Vzhledem k neporovnatelnosti hodnot věrohodností klasického logistického modelu a logistického modelu s náhodnými efekty, musíme signifikanci náhodného efektu ověřit jinými prostředky. Můžeme např. spočítat interval spolehlivosti pro parametr σ^2.

Interval spolehlivosti pro σ^2 můžeme získat např. pomocí profilování věrohodnosti. Hlavní myšlenka této metody vychází z testu poměru věrohodností. Výsledný $100(1-\alpha)\%$ interval spolehlivosti pro σ^2 je množina všech hodnot σ_0^2

takových, že oboustranný test nulové hypotézy $\mathrm{H}_0\colon \sigma^2 = \sigma_0^2$ není zamítnut na hladině významnosti α. Testová statistika takového testu je pro pevnou hodnotu σ_0^2

$$G = -2\ln\left[\frac{L(\sigma_0^2, \hat{\boldsymbol{\beta}}_0)}{L(\hat{\sigma}^2, \hat{\boldsymbol{\beta}})}\right] = -2\left[l(\sigma_0^2, \hat{\boldsymbol{\beta}}_0) - l(\hat{\sigma}^2, \hat{\boldsymbol{\beta}})\right],$$

kde $L(\sigma_0^2, \hat{\boldsymbol{\beta}}_0)$ je profilová věrohodnostní funkce pro σ_0^2

$$L(\sigma_0^2, \hat{\boldsymbol{\beta}}_0) = \max_{\boldsymbol{\beta} \in \mathbb{R}^{k+1}} L(\sigma_0^2, \boldsymbol{\beta})$$

a $\hat{\sigma}^2, \hat{\boldsymbol{\beta}}$ jsou odhady získané metodou maximální věrohodnosti. Na základě asymptotického χ^2 rozdělení s 1 stupněm volnosti testové statistiky G můžeme psát přibližný interval spolehlivosti pro σ^2 jako

$$\mathcal{I}_{1-\alpha}(\sigma^2) = \left\{\sigma_0^2 : -2\left[l(\sigma_0^2, \hat{\boldsymbol{\beta}}_0) - l(\hat{\sigma}^2, \hat{\boldsymbol{\beta}})\right] \leq \chi_1^2(1-\alpha)\right\}.$$

Pokud interval spolehlivosti neobsahuje nulu, řekneme, že náhodný efekt je signifikantní.

Výběr a hodnocení logistického modelu s náhodným absolutním členem můžeme provést stejným způsobem jako u klasického logistického modelu (viz podkapitola 1.5). Vhodné proměnné do modelu lze vybrat na základě krokové selekce a Akaikeho informačního kritéria. Dodržení předpokladů modelu můžeme také zkontrolovat stejným způsobem, včetně ověření multikolinearity pomocí hodnot VIF. A díky tomu, jak je počítána testová statistika Hosmerova-Lemeshowova testu dobré shody, je tento test aplikovatelný i na logistický model s náhodným absolutním členem.

Kapitola 3

Představení dat

V rámci této kapitoly si představíme data, jejichž analýzu si ukážeme v kapitole následující. Data byla poskytnuta *Ústavem molekulární a translační medicíny* v Olomouci a obsahují informace o 170 pacientech s nitrolebním nádorem, tzv. meningeomem. Výsledky studie, ze které data pochází, byly publikovány v článku *Identification of Meningioma Patients at High Risk of Tumor Recurrence Using MicroRNA Profiling* [18].

Meningeom je nádorové onemocnění mozkového obalu, konkrétně se jedná o nádor pavučnice, což je prostřední ze tří blan nacházejících se mezi lebeční kostí a mozkem. Meningeomy představují přibližně 15 až 20 % všech nitrolebních nádorů. Často je jejich růst velmi pomalý, tudíž se u pacienta mohou vyskytovat nepozorovaně i několik let. [11]

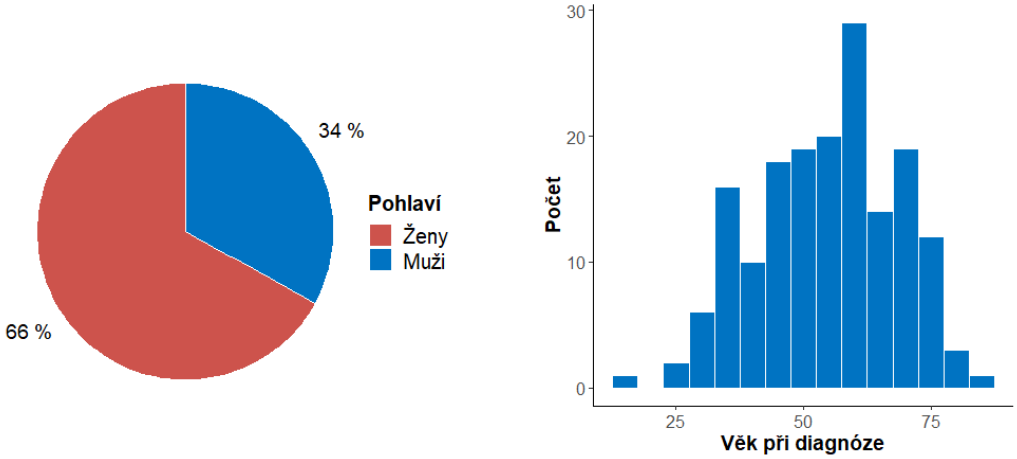
Datový soubor sestává z pozorování o 170 pacientech, o kterých máme k dispozici základní údaje, jako je pohlaví a věk při diagnóze, dále pak specifické charakteristiky meningeomu a nakonec také výsledky kvantitativní PCR metody. Pojďme si nyní představit jednotlivé proměnné trochu podrobněji spolu se stručnou popisnou statistikou.

Pohlaví

Kvalitativní veličina s kategoriemi *muž* a *žena*. Datový soubor je v tomto ohledu poněkud nevyrovnaný. Mezi 170 pacienty je 113 žen a pouze 57 mužů, což reflektuje skutečnost, že ženy trpí tímto onemocněním častěji než muži (Obrázek 3.1).

Věk při diagnóze

Kvantitativní veličina, jejíž histogram vidíme na Obrázku 3.1. Nejmladšímu pacientu byl meningeom diagnostikován v 17 letech, nejstaršímu v 86 letech. Průměrným věkem při diagnóze je necelých 55 let.



Obrázek 3.1: Koláčový graf proměnné Pohlaví (vlevo) a histogram proměnné Věk při diagnóze (vpravo)

Grading meningeomu

Grading meningeomu je kvalitativní proměnná s kategoriemi *WHO grade I*, *WHO grade II* a *WHO grade III*. Jedná se o ustálenou klasifikaci meningeomů zavedenou Mezinárodní zdravotnickou organizací (WHO - *World Health Organization*) do tří skupin následovně

- WHO grade I - pomalu rostoucí nádor, často označován jako nezhoubný (benigní),
- WHO grade II - rychleji rostoucí atypický nádor,
- WHO grade III - rychle rostoucí nádor, často označován jako zhoubný (maligní). [12]

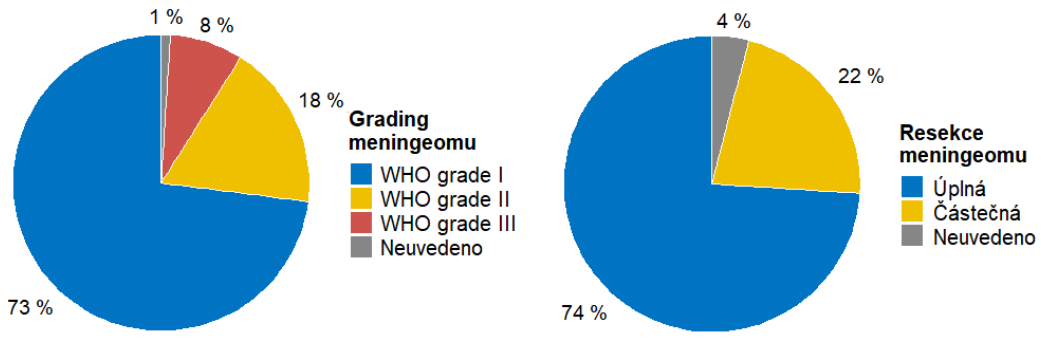
Zastoupení jednotlivých kategorií v našem datovém souboru opět odpovídá známé skutečnosti, že nejhojněji zastoupenou skupinou jsou nádory klasifikované

jako WHO grade I. Ty se objevují u 124 pacientů. Nádor typu WHO grade II pozorujeme u 31 pacientů, nádor typu WHO grade III u 13 pacientů a u 2 pacientů nebyl grading nádoru uveden (Obrázek 3.2).

Resekce meningeomu

Kvalitativní proměnná s kategoriemi *úplná* a *částečná*. Meningeom je nádorové onemocnění, které je možné léčit radioterapií, výjimečně chemoterapií, v naprosté většině případů je ale léčeno resekcí, tedy operativním odstraněním celého nádoru, či jeho části. Pokud to lokace a charakter nádoru umožňuje, je preferována úplná resekce před částečnou. [2]

Studie, ze které naše data pochází, byla zaměřena na pacienty, kteří podstoupili resekci meningeomu. Ze 170 pacientů podstoupilo 127 pacientů úplnou resekci, 37 pacientů částečnou resekci a u 6 pacientů nebyl rozsah resekce uveden (Obrázek 3.2).



Obrázek 3.2: Koláčový graf proměnné Grading meningeomu (vlevo) a proměnné Resekce meningeomu (vpravo)

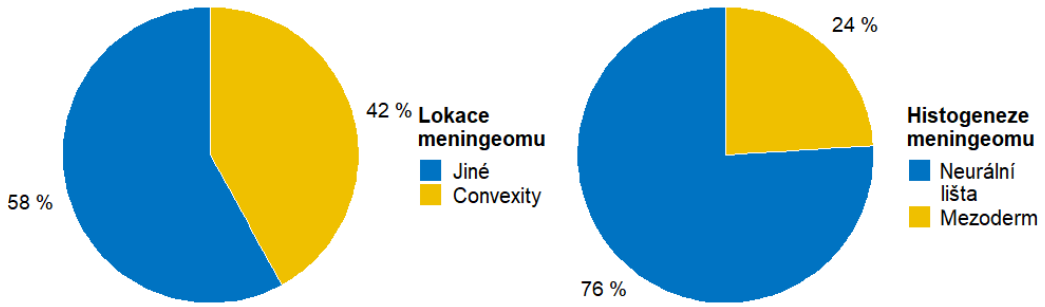
Lokace meningeomu

Kvalitativní proměnná s kategoriemi *convexity* a *jiné*. Meningeomy označené kategorií *convexity* rostou na vnější části mozku, jsou tedy snadno přístupné. V našem datovém souboru se vyskytují u 71 pacientů. Jiné umístění meningeomu může být například za okem, za nosními dutinami či v blízkosti mozkového

kmene. Takto umístěný meningeom pozorujeme u 99 pacientů (Obrázek 3.3).

Histogeneze meningeomu

Meningeomy se dělí také na základě histogeneze, tedy podle tkáně, jejíž rysy alespoň částečně přebírá i tkáň nádorová a ze které pravděpodobně nádor vyrostl [21]. V datovém souboru tato kvalitativní proměnná sestává z kategorie *neurální lišta* se zastoupením 129 pacientů a z kategorie *mezoderm* se zastoupením 41 pacientů (Obrázek 3.3).



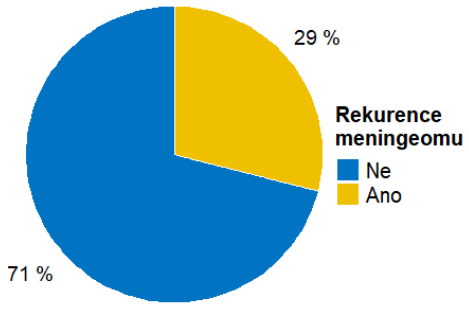
Obrázek 3.3: Koláčový graf proměnné Lokace meningeomu (vlevo) a proměnné Histogeneze meningeomu (vpravo)

Rekurence meningeomu

Stejně jako ve studii, ze které naše data pochází, i naším cílem bude především zkoumat, jaké charakteristiky nám pomohou identifikovat pacienty, u kterých je vyšší riziko rekurence meningeomu. Rekurenci meningeomu nelze spolehlivě předvídat pouze na základě zhoubnosti nádoru. U benigních nádorů (WHO grade I) dochází k rekurenci do 5 let po úplné resekci u 12 % všech případů, po částečné resekci dokonce až u 37-60 % případů. [18]

Naší vysvětlovanou proměnnou v následné analýze dat bude právě rekurence meningeomu, konkrétně rekurence do 8 let, tj. do 96 měsíců, od poslední resekce. Tuto binární proměnnou jsme odvodili na základě proměnné TTR (*Time to relaps*) měřené v měsících, jejíž závislost na ostatních proměnných modelovali

autoři článku [18] pomocí analýzy přežití. V našem datovém souboru došlo k rekurenci do 8 let od resekce u 49 pacientů, u 121 pacientů k ní nedošlo (Obrázek 3.4).



Obrázek 3.4: Koláčový graf proměnné Rekurence meningeomu

MicroRNA

MicroRNA, zkráceně miRNA, je krátká nekódující RNA podílející se na regulaci genové exprese ostatních typů RNA, zvláště pak mRNA. Bylo dokázáno, že některé miRNA mohou hrát roli onkogenů, či naopak supresorů u některých nádorových onemocnění. Stále častěji je proto analýza profilů miRNA využívána za účelem nalezení nových biomarkerů pro diagnostiku těchto onemocnění. Výhoda miRNA jako biomarkeru je taková, že je velmi stabilní, a to často i v tělních tekutinách, jako jsou sliny či moč. Získání vzorku pro analýzu je tím pádem mnohem méně invazivní než například odebírání tkáně. [15], [18]

V rámci studie, ze které naše data pochází, bylo u každého pacienta studováno několik různých miRNA. Cílem studie bylo zjistit, zda některá ze studovaných miRNA hraje roli identifikátoru pacientů s vysokým rizikem rekurence meningeomu. Včasné zachycení těchto pacientů by pak mohlo vést k cílené, potenciálně i účinnější léčbě.

V naší analýze se konkrétně zaměříme na sedm různých miRNA, a to *miR-331-3p*, *miR-18a-5p*, *miR-16-5p*, *miR-15a-5p*, *miR-146a-5p*, *miR-130b-3p* a *miR-*

1271-5p. Každému pacientovi byl odebrán vzorek tkáně, ze kterého byla izolována celková RNA. Exprese jednotlivých miRNA byla následně změřena pomocí kvantitativní PCR metody (viz níže). Měření proběhlo ve třech technických replikátech, jinými slovy u každého pacienta byla k dispozici tři měření pro každou miRNA. Následně byla z datového souboru vyřazena odlehlá pozorování, u kterých existovalo na základě přezkoumání důvodné podezření, že došlo k technické chybě při měření. V konečném důsledku tedy máme v datech 3214 pozorování, tj. u každého ze 170 pacientů máme k dispozici 1-3 pozorování pro všechny výše uvedené miRNA.

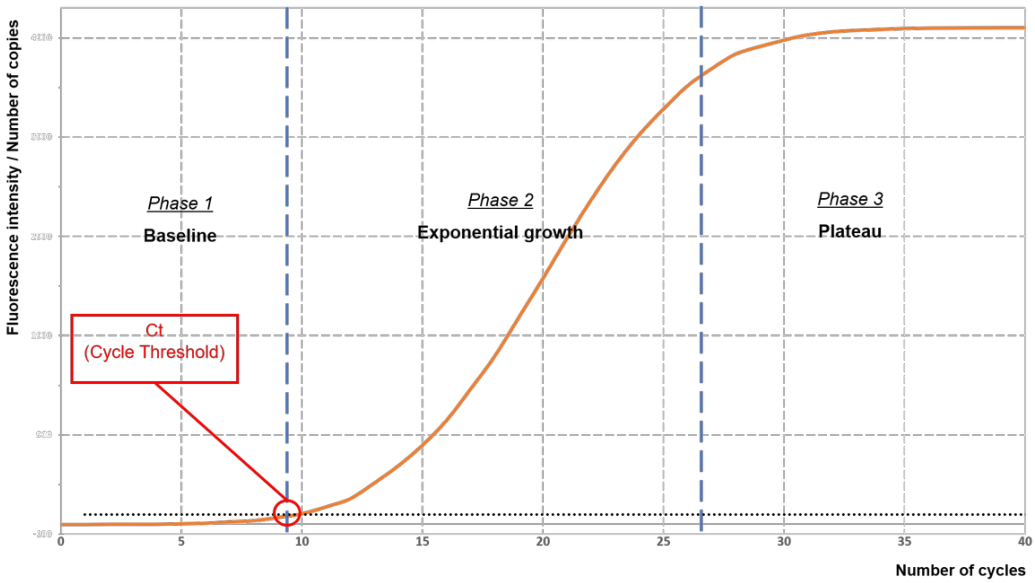
Mimo výše uvedená pozorování máme k dispozici navíc 509 pozorování pro *miR-181b-5p*. Tuto informaci uvádíme zvlášť, jelikož tato miRNA je dále použita pouze jako norma (viz níže), tudíž ji v další analýze nebudeme uvažovat jako potenciální biomarker rekurence meningeomu.

Výsledky kvantitativní PCR metody

Kvantitativní polymerázová řetězová reakce v reálném čase (RT qPCR - *Real Time Quantitative Polymerase Chain Reaction*) je laboratorní technika molekulární biologie, pomocí níž sledujeme amplifikaci (znásobení množství) určitého vzorku DNA. Při qPCR založené na RNA je třeba nejprve vzorek RNA přetransformovat na komplementární DNA (cDNA - *Complementary DNA*) pomocí tzv. reverzní transkriptázy. Takto získané vzorky DNA jsou následně smíchány s řadou dalších látek včetně fluorescenčního barviva. Směs je poté v přístroji, který nazýváme termocykler, cyklicky zahřívána a ochlazována, čímž dojde k amplifikaci DNA. Rychlost růstu množství DNA je sledována pomocí měření fluorescence. [22]

Výstupem qPCR v reálném čase je tzv. amplifikační křivka (Obrázek 3.5). Amplifikační křivka zobrazuje vývoj míry fluorescence v závislosti na počtu provedených cyklů a skládá se ze tří fází. První fáze je inicializační (na Obrázku 3.5 označená jako *baseline*), při které míra fluorescence výrazně neroste. Druhá fáze, při níž dochází k rychlému růstu míry fluorescence, se nazývá fáze exponenciálního růstu. Poslední třetí fáze se označuje jako plateau fáze, kdy růst míry

fluorescence zpomaluje, až se stabilizuje.

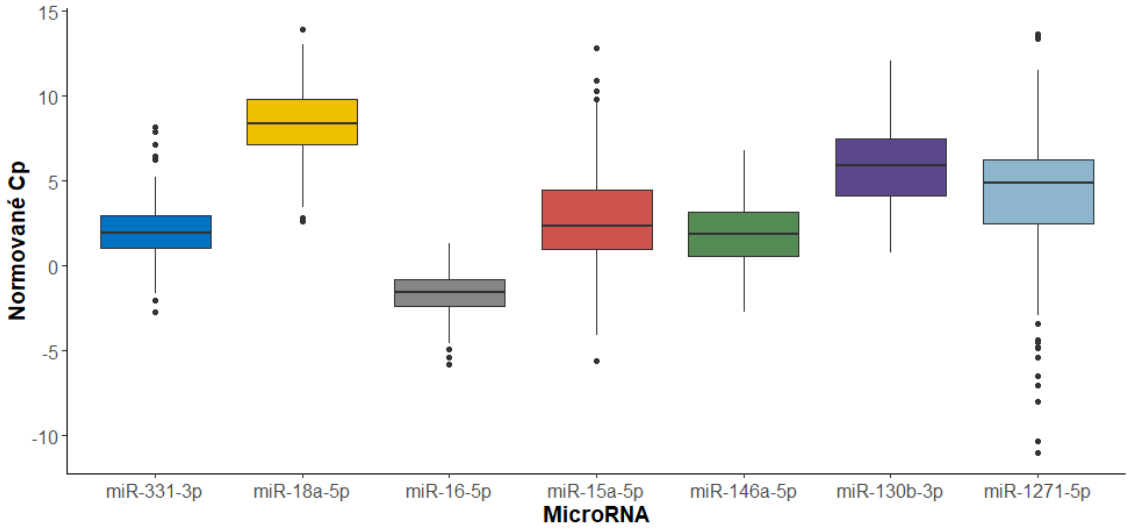


Obrázek 3.5: Schéma amplifikační křivky qPCR metody v reálném čase
Zdroj: [1]

Hlavní výstupní hodnotou tohoto grafu je tzv. *crossing point* (Cp), někdy též *cycle threshold* (Ct). Jedná se o počet cyklů, ve kterém amplifikační křivka dosáhne prahové hodnoty (*thresholdu*), která je předem stanovena. Čím menší hodnota Cp, tím větší množství RNA, resp. cDNA, bylo na počátku řetězové reakce.

K porovnatelnosti výsledných hodnot Cp mezi jednotlivými pacienty slouží tzv. normování. Od hodnoty Cp pro námi sledované miRNA odečteme tzv. normu. Norma je hodnota Cp pro miRNA, jejíž exprese je stabilní, vysoká a nemění se v závislosti na sledovaných parametrech.

V našich datech máme celkem 3214 napozorovaných hodnot Cp pro výše uvedené miRNA. Jako norma byl u každého pacienta použit aritmetický průměr hodnot Cp pro *miR-181b-5p*. Po znormování jsme dostali hodnoty Cp, jejichž krabicové grafy vidíme na Obrázku 3.6.



Obrázek 3.6: Krabicové grafy proměnné Normované Cp v závislosti na MicroRNA

Kapitola 4

Analýza dat

V rámci této kapitoly si představíme postup a výsledky analýzy dat o 170 pacientech, se kterými jsme se seznámili v předešlé kapitole. Nejprve se zaměříme na klasickou logistickou regresi, následně na logistický model s náhodným absolutním členem. Veškerá analýza je provedena ve statistickém softwaru R verze 4.1.2.

4.1. Vícenásobná logistická regrese

Naším cílem je pomocí nástrojů logistické regrese zjistit, které faktory mají vliv na rekurenci meningeomu do 8 let od poslední resekce. Proměnná Rekurence meningeomu bude tedy v roli vysvětlované proměnné. Vysvětlujícími proměnnými pak budou proměnné Pohlaví, Věk při diagnóze, Grading meningeomu, Resekce meningeomu, Lokace meningeomu, Histogeneze meningeomu a především výsledek qPCR metody, Normované Cp.

Pro každou miRNA, jejíž exprese byla u pacientů zkoumána, sestavíme logistický model, celkem tedy dostaneme 7 modelů. Abychom se vypořádali s opakovanými měřeními hodnot Cp, u každého pacienta vždy vezmeme aritmetický průměr těchto hodnot pro danou miRNA. Kvůli odlišnému měřítku kvantitativních proměnných Věk při diagnóze a Normované Cp navíc tyto dvě proměnné standardizujeme, tj. od napozorovaných hodnot proměnných odečteme jejich aritmetický průměr a podělíme jejich směrodatnou odchylkou. V softwaru R k tomuto

účelu slouží funkce `scale`. Hodnoty aritmetických průměrů a směrodatných odchylek proměnných Věk při diagnóze a Normované Cp před standardizací pro jednotlivé modely nalezneme v Tabulce 4.1.

	Věk při diagnóze		Normované Cp	
Model	Průměr	Sm. odchylka	Průměr	Sm. odchylka
miR-331-3p	54.29	13.79	2.00	1.48
miR-18a-5p	54.32	13.81	8.53	1.90
miR-16-5p	54.21	13.79	-1.67	1.11
miR-15a-5p	54.11	13.84	2.91	2.73
miR-146a-5p	54.21	13.79	1.95	1.80
miR-130b-3p	54.45	13.67	5.90	2.26
miR-1271-5p	54.31	13.40	3.95	3.45

Tabulka 4.1: Hodnoty aritmetických průměrů a směrodatných odchylek proměnných Věk při diagnóze a Normované Cp před standardizací

Pro jednoduchost uvedeme nejprve postup a komentář výsledků analýzy pouze pro jeden z modelů, a to pro *miR-331-3p*. Jako první sestavíme úplný model se všemi vysvětlujícími proměnnými, které máme k dispozici, tedy

$$\textit{Rekurence} \sim \textit{Věk při diagn.} + \textit{Pohlaví} + \textit{Grading} + \textit{Resekce} + \textit{Lokace} + \textit{Histogeneze} + \textit{Normované Cp}.$$

Výstup funkce `glm` získaný pomocí funkce `summary` vidíme v Tabulce 4.2. V prvním sloupci jsou uvedené názvy proměnných. Ve druhém a třetím sloupci uvádíme odhady regresních parametrů β_j spolu s jejich 95% intervaly spolehlivosti (CI - *Confidence Interval*). V dalších dvou sloupcích jsou napsané odhady poměrů šancí (OR), tedy $\exp(\hat{\beta}_j)$, a jejich intervaly spolehlivosti. V posledním sloupci jsou pak uvedené p-hodnoty testů významnosti regresních parametrů, p-hodnoty menší než hladina významnosti 0.05 jsou označené červenou barvou.

Vidíme, že ne všechny proměnné jsou signifikantní, pro nás nejdůležitější proměnná Normované Cp ale signifikantní je. Bodový odhad poměru šancí je roven hodnotě 4.20, znamená to tedy, že zvýší-li se hodnota Cp o jednotku, šance

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-3.21	(-4.67; -1.96)	-	-	< 0.0001
Věk při diagnóze	-0.29	(-0.75; 0.16)	0.75	(0.47; 1.17)	0.2046
Pohlaví - Žena	0.27	(-0.67; 1.26)	1.31	(0.51; 3.53)	0.5761
Grading - II	0.89	(-0.13; 1.91)	2.44	(0.88; 6.76)	0.0843
Grading - III	2.41	(0.73; 4.17)	11.12	(2.07; 64.64)	0.0052
Resekce - Částečná	1.53	(0.55; 2.55)	4.60	(1.73; 12.84)	0.0026
Lokace - Convexity	1.53	(0.46; 2.71)	4.61	(1.58; 14.99)	0.0072
Histog. - Mezoderm	1.04	(-0.19; 2.31)	2.82	(0.83; 10.04)	0.1006
Normované Cp	1.43	(0.87; 2.09)	4.20	(2.38; 8.12)	< 0.0001

Tabulka 4.2: Tabulka výsledků úplného logistického modelu pro *miR-331-3p*

rekurence meningeomu bude v průměru 4.20krát vyšší. Vzhledem k tomu, že jsme tuto kvantitativní proměnnou standardizovali, změna o jednotku odpovídá změně o jednu směrodatnou odchylku, tedy přibližně o 1.48 jednotek původní nestandardizované proměnné Normované Cp (viz Tabulka 4.1).

Další signifikantní proměnnou je Resekce meningeomu, bodový odhad poměru šancí pro tuto proměnnou je 4.60, šance rekurence meningeomu je tedy 4.60krát vyšší pro pacienty s částečnou resekcí, než pro pacienty s úplnou resekcí meningeomu. Podobně pro pacienty s lokací meningeomu označenou jako Convexity je šance rekurence 4.61krát vyšší, než pro pacienty s jinou lokací meningeomu. Nakonec pro pacienty se zhoubnou formou meningeomu (WHO grade III) je 11.12krát vyšší šance rekurence meningeomu, než pro pacienty s nezhoubnou formou (WHO grade I). Můžeme si všimnout, že tento konkrétní bodový odhad poměru šancí není moc přesný. Jeho interval spolehlivosti (2.07; 64.64) je dosti široký. Příčinou může být malý počet pozorování u kategorie WHO grade III.

Nemůžeme opomenout ani interpretaci absolutního členu. Hodnota exponenciální transformace bodového odhadu absolutního členu odpovídá šanci rekurence meningeomu u pacienta s referenčními kategoriemi kategoriálních proměnných a nulovými hodnotami kvantitativních proměnných. Uvědomme si, že nulová hodnota standardizované kvantitativní proměnné odpovídá průměrné hodnotě

původní nestandardizované proměnné. Celkově proto řekneme, že šance rekurence meningeomu pro 54letého muže, jemuž byl odoperován celý nezhoubný meningeom s jinou než convexikální lokací, s kategorií histogeneze Neurální lišta a s Normovaným Cp o hodnotě 2, je rovna hodnotě $\exp(-3.21) = 0.0404$. Převedeme-li tuto hodnotu do tvaru, který se u pojmu šance používá, šance rekurence meningeomu pro takového muže je zhruba 1:25.

Vzhledem k tomu, že jsme do úplného modelu zahrnuli i proměnné, které nejsou signifikantní, můžeme zkusit vytvořit podmodel úplného modelu. Pro výběr proměnných, které v modelu ponecháme, použijeme sestupnou krokovou selekci. Výsledným zjednodušeným modelem je model

$$\textit{Rekurence} \sim \textit{Grading} + \textit{Resekce} + \textit{Lokace} + \textit{Histogeneze} + \textit{Normované Cp}.$$

Vynechanými proměnnými oproti úplnému modelu jsou tedy proměnné Pohlaví a Věk při diagnóze. Výsledky zjednodušeného modelu můžeme vidět v Tabulce 4.3.

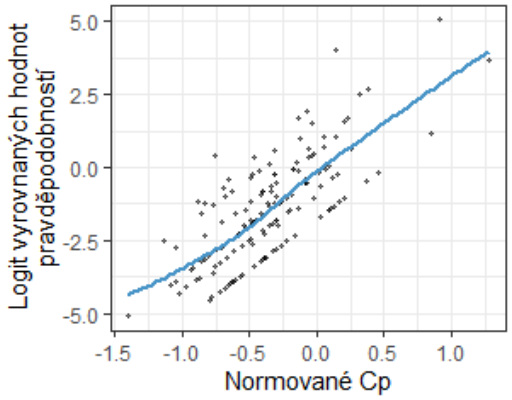
Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-2.95	(-4.12; -1.98)	-	-	< 0.0001
Grading - II	0.80	(-0.21; 1.81)	2.23	(0.81; 6.11)	0.1174
Grading - III	2.13	(0.54; 3.73)	8.40	(1.72; 41.75)	0.0077
Resekce - Částečná	1.63	(0.67; 2.64)	5.08	(1.95; 13.99)	0.0011
Lokace - Convexity	1.50	(0.43; 2.68)	4.47	(1.53; 14.52)	0.0085
Histog. - Mezoderm	1.00	(-0.19; 2.24)	2.72	(0.83; 9.36)	0.1034
Normované Cp	1.37	(0.84; 1.99)	3.93	(2.32; 7.29)	< 0.0001

Tabulka 4.3: Tabulka výsledků zjednodušeného logistického modelu pro *miR-331-3p*

Vidíme, že výsledky se zásadně nezměnily, odhady jsou ale díky vyřazení přebytečných proměnných přesnější. Zkusíme nyní provést test podmodelu (test poměru věrohodností). V softwaru R lze pro tento test jednoduše použít funkci `anova`. Výsledná p-hodnota je rovna 0.3446, a jelikož překračuje hladinu významnosti 0.05, nezamítáme nulovou hypotézu o nulovosti regresních parametrů od-

povídajících proměnným Pohlaví a Věk při diagnóze. Dále tedy budeme uvažovat zjednodušený model.

Zkusme nyní ověřit splnění předpokladů tohoto modelu. Jedním z předpokladů je lineární vztah mezi kvantitativními vysvětlujícími proměnnými a logitem pravděpodobnosti nastání jevu daného vysvětlovanou proměnnou, $g(\boldsymbol{x}_i)$, $i = 1, \ldots, n$. V našem případě se jedná o vztah mezi hodnotami proměnné Normované Cp a logitem pravděpodobnosti rekurence meningeomu do 8 let od poslední resekce. Pomocí příkazu `predict(glm_objekt,type=response)` získáme vyrovnané hodnoty pravděpodobnosti rekurence meningeomu $\hat{\pi}(\boldsymbol{x}_i)$, na které následně aplikujeme logitovou transformaci, čímž získáme $\hat{g}(\boldsymbol{x}_i)$. Nyní stačí vykreslit graf zmíněné závislosti (Obrázek 4.1). Pro lepší představu o tvaru závislosti jsme využili funkci `geom_smooth` z knihovny `ggplot2`, jejímž výsledkem je modrá křivka v grafu. Vidíme, že vztah je na první pohled lineární, takže tento předpoklad je splněn.



Obrázek 4.1: Graf vztahu logitu vyrovnaných hodnot pravděpodobností a hodnot proměnné Normované Cp zjednodušeného logistického modelu pro *miR-331-3p*

Nyní zkusíme spočítat VIF pro ověření, zda se v datech nevyskytuje multikolinearita. V programu R lze pro výpočet snadno použít funkci `vif` z knihovny `car`, jejíž výstup vidíme v Tabulce 4.4.

Veškeré hodnoty VIF se pohybují velmi blízko jedné, můžeme tedy říct, že multikolinearitu v datech nepozorujeme. Nakonec provedeme Hosmerův-Lemeshowův test pomocí funkce `hosmerlem_test` z knihovny `VADIS`. Výsledná p-hodnota je rovna hodnotě 0.7253, nezamítáme tedy nulovou hypotézu o správnosti modelu.

Celkově můžeme konstatovat, že máme dobrý model se splněnými předpoklady.

Proměnná	VIF
Grading	1.0456
Resekce	1.0620
Lokace	1.3210
Histogeneze	1.2184
Normované Cp	1.1220

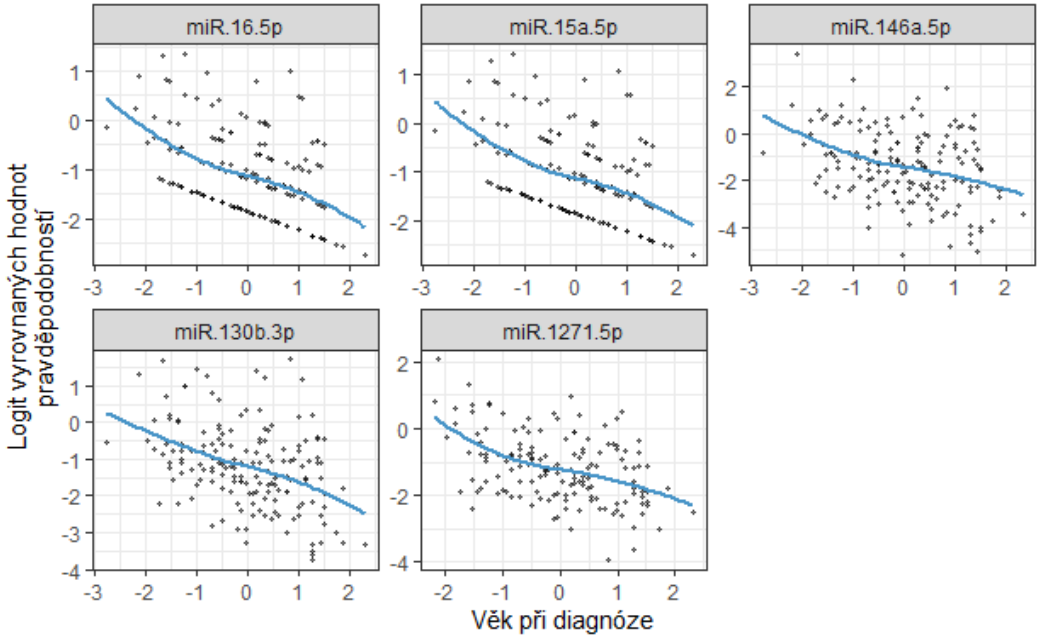
Tabulka 4.4: Hodnoty VIF pro proměnné zjednodušeného logistického modelu pro *miR-331-3p*

Stejný postup jsme aplikovali i pro zbylých 6 modelů. Výsledky úplných modelů nalezneme v příloze A v Tabulkách A.1 až A.6. Pro nás nejdůležitější proměnná Normované Cp je statisticky signifikantní celkem ve třech modelech ze šesti, a to pro *miR-18a-5p*, *miR-146a-5p* a *miR-130b-3p*. U jednoho dalšího modelu jsme pozorovali p-hodnotu testu významnosti u proměnné Normované Cp menší než 0.1, a to pro *miR-1271-5p*. Pro běžně volenou hladinu významnosti 0.05 tedy nulovost regresního parametru nezamítáme jen těsně. Proměnnou, která je signifikantní ve všech sedmi modelech, je Resekce meningeomu. Pro pacienty s částečnou resekcí je šance rekurence signifikantně vyšší než pro pacienty s úplnou resekcí, což je očekávatelný výsledek. Ve čtyřech modelech ze šesti je také signifikantní proměnná Lokace meningeomu.

Nyní opět použijme sestupnou krokovou selekci pro zjednodušení modelů. Výsledky zjednodušených modelů nalezneme v příloze B v Tabulkách B.1 až B.6, v popisku tabulek pak nalezneme p-hodnotu testu podmodelu. U všech testů podmodelu byla p-hodnota větší než hladina významnosti 0.05, nezamítáme tedy nulovou hypotézu o nulovosti regresních parametrů a dále budeme uvažovat zjednodušené modely. Většinou byla z úplných modelů vyřazena proměnná Pohlaví, a to ze třech modelů z šesti. Ze dvou jiných modelů z šesti pak byly vyřazeny proměnné Lokace, Histogeneze a Normované Cp. Proměnná Normované Cp tedy zůstala zahrnuta celkem ve čtyřech modelech, a to ve všech jako statisticky signifikantní proměnná. Proměnná Resekce meningeomu zůstala sig-

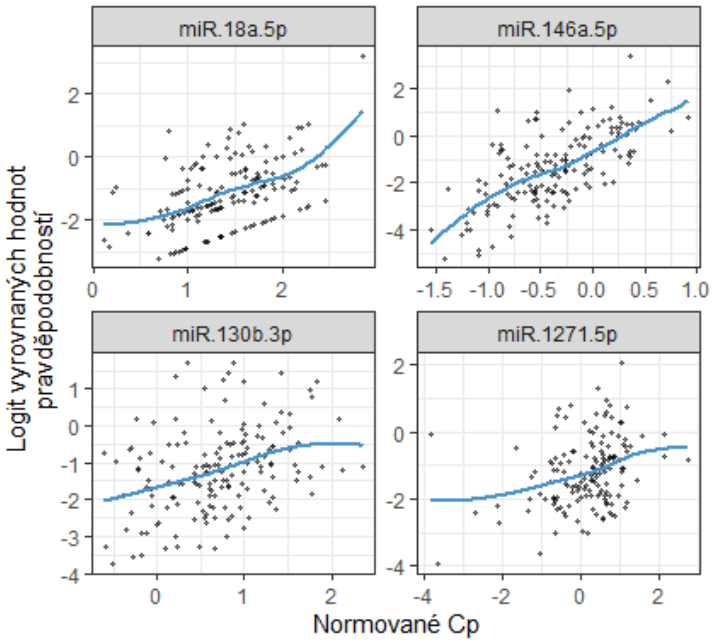
nifikantní proměnnou ve všech šesti modelech.

Nyní opět ověříme splnění předpokladů modelů. Nejprve se zaměříme na lineární vztah mezi kvantitativními vysvětlujícími proměnnými a logitem pravděpodobnosti rekurence meningeomu. Na Obrázku 4.2 vidíme grafy pro proměnnou Věk při diagnóze, která byla zahrnuta ve všech zjednodušených modelech, kromě modelu pro *miR-18a-5p*. Vztah se pro všechny modely zdá být lineární.



Obrázek 4.2: Graf vztahu logitu vyrovnaných hodnot pravděpodobnosti a hodnot proměnné Věk při diagnóze zjednodušených logistických modelů

V grafech na Obrázku 4.3 vidíme obdobný vztah pro proměnnou Normované Cp. Tato proměnná byla zařazena pouze do čtyř zjednodušených modelů. Zejména u modelu *miR-18a-5p* bychom o linearitě vztahu mohli trochu pochybovat. Proto jsme do modelů zkusili zahrnout druhou mocninu proměnné Normované Cp, v žádném z modelů ale nevyšla signifikantně (nejmenší p-hodnota testu významnosti byla právě v modelu pro *miR-18a-5p* a byla rovna 0.2121). Podotkněme nicméně, že druhé mocniny, a obecně polynomy, nejsou jedinou možností, jak modelovat nelineární vztah proměnných. V rámci této diplomové práce se ale jinými možnými transformacemi zabývat nebudeme.



Obrázek 4.3: Graf vztahu logitu vyrovnaných hodnot pravděpodobnosti a hodnot proměnné Normované Cp zjednodušených logistických modelů

Podívejme se ještě na hodnoty VIF pro kontrolu, zda se nepotýkáme s multikolinearitou v datech. V Tabulce 4.5 vidíme, že všechny hodnoty VIF pro všech šest zbylých modelů jsou blízké jedné, multikolinearitu tedy v datech nemáme.

Proměnná	18a-5p	16-5p	15a-5p	146a-5p	130b-3p	1271-5p
Věk při diagn.	-	1.0323	1.0316	1.0460	1.0377	1.0310
Pohlaví	-	1.0245	1.0269	-	1.0303	-
Grading	1.0182	1.0149	1.0146	1.0756	1.0538	1.0738
Resekce	1.0401	1.0075	1.0082	1.0259	1.0369	1.0490
Lokace	1.2692	-	-	1.2609	1.0817	1.2783
Histogeneze	1.2509	-	-	1.2415	-	1.2657
Norm. Cp	1.0230	-	-	1.1407	1.0964	1.1702

Tabulka 4.5: Hodnoty VIF pro proměnné zjednodušených logistických modelů (V prvním řádku tabulky jsou označení miRNA uvedeny bez předpony *miR*)

Nakonec se ještě podívejme na výsledky Hosmerova-Lemeshowova testu dobré shody. Výsledné p-hodnoty jsou uvedeny v Tabulce 4.6. Všechny p-hodnoty jsou

větší než 0.05, nulovou hypotézu o správnosti modelu proto nezamítáme ani u jednoho zjednodušeného modelu.

Model	p-hodnota
miR-18a-5p	0.2710
miR-16-5p	0.1144
miR-15a-5p	0.2784
miR-146a-5p	0.9952
miR-130b-3p	0.4742
miR-1271-5p	0.7207

Tabulka 4.6: Výsledné p-hodnoty Hosmerova-Lemeshowova testu dobré shody zjednodušených logistických modelů

Na základě klasické logistické regrese můžeme celkově konstatovat, že jako biomarkery rekurence meningeomu by potenciálně mohly sloužit *miR-331-3p* a *miR-146a-5p*. U *miR-18a-5p*, *miR-130b-3p* a *miR-1271-5p* jsme nulovou hypotézu testu významnosti zamítali jen těsně. Také jsme analýzou potvrdili všeobecně známý fakt, že úplná resekce by měla být upřednostňována před částečnou, pokud je to možné. Pacienti s částečnou resekcí mají signifikantně vyšší šanci rekurence meningeomu.

4.2. Vícenásobná logistická regrese s náhodným absolutním členem

Nyní do výše sestavených modelů přidáme náhodný absolutní člen a výsledky porovnáme s těmi získanými klasickou logistickou regresí. Znovu chceme zjistit, které proměnné mají signifikantní vliv na binární závisle proměnnou Rekurence meningeomu do 8 let od resekce. Náhodným efektem je v našich datech kategoriální proměnná Pacient, která odpovídá anonymizovanému označení jednotlivých pacientů. Opakovaná měření proměnné Normované Cp jsou pro jednotlivé pacienty závislá, díky zahrnutí náhodného efektu do modelu je ale nemusíme agregovat pomocí aritmetického průměru, jako tomu bylo u klasického logistického

modelu. Kvantitativní proměnné Věk při diagnóze a Normované Cp opět standardizujeme pomocí aritmetického průměru a směrodatné odchylky, které nalezneme v Tabulce 4.7.

	Věk při diagnóze		Normované Cp	
Model	Průměr	Sm. odchylka	Průměr	Sm. odchylka
miR-331-3p	53.96	13.73	1.99	1.47
miR-18a-5p	54.06	13.57	8.49	1.91
miR-16-5p	54.07	13.71	-1.67	1.14
miR-15a-5p	54.07	13.83	2.81	2.71
miR-146a-5p	54.15	13.72	1.95	1.80
miR-130b-3p	54.54	13.59	5.85	2.24
miR-1271-5p	53.66	13.47	4.37	3.24

Tabulka 4.7: Hodnoty aritmetických průměrů a směrodatných odchylek proměnných Věk při diagnóze a Normované Cp před standardizací

Sestavme nejprve úplný logistický model s náhodným absolutním členem

$$\textit{Rekurence} \sim \textit{Věk při diagn.} + \textit{Pohlaví} + \textit{Grading} + \textit{Resekce} + \textit{Lokace} + \textit{Histogeneze} + \textit{Normované Cp} + (1|\textit{Pacient}).$$

Náhodný efekt ve formě absolutního členu zapisujeme jako $(1|Pacient)$. Stejný zápis je vyžadován i v rámci funkce `glmer` z knihovny `lme4`, kterou použijeme pro odhadování parametrů. Kvůli prvotním problémům s konvergencí jsme nastavili parametr `control = glmerControl(optimizer="bobyqa")`. Tím jsme změnili optimalizační metodu, která je v rámci funkce `glmer` použita pro nalezení optimálního řešení, z defaultního Nelder-Meadova algoritmu na omezenou optimalizaci pomocí kvadratické aproximace (*Bound optimization by quadratic approximation*). Stejně jako v předchozí podkapitole uvedeme výsledky a komentář nejprve pro *miR-331-3p*. V Tabulce 4.8 nalezneme výstup funkce `glmer` pro tuto miRNA.

První skutečností, které si můžeme všimnout, je, že jsme sloučili kategorie WHO grade II a WHO grade III proměnné Grading meningeomu. Sloučili jsme

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodn.
Absolutní člen	-10.50	(-14.88; -6.12)	-	-	< 0.0001
Věk při diagn.	-1.21	(-2.57; 0.16)	0.30	(0.08; 1.17)	0.0828
Pohlaví - Muž	0.36	(-2.29; 3.01)	1.43	(0.10; 20.31)	0.7899
Grading - II-III	4.61	(1.54; 7.68)	100.24	(4.65; 2162.33)	0.0033
Resekce - Část.	6.10	(2.57; 9.63)	444.35	(13.03; 15156.25)	0.0007
Lokace - Conv.	4.24	(0.94; 7.54)	69.35	(2.56; 1880.78)	0.0118
Histog. - Mez.	3.53	(-0.07; 7.12)	34.02	(0.93; 1238.91)	0.0545
Normované Cp	3.13	(1.77; 4.48)	22.86	(5.90; 88.57)	< 0.0001
Parametr σ	5.01	(3.83; 6.89)	-	-	-

Tabulka 4.8: Tabulka výsledků úplného logistického modelu s náhodným absolutním členem pro *miR-331-3p*

je kvůli velmi širokému intervalu spolehlivosti u kategorie WHO grade III (v porovnání s ostatními), který byl zapříčiněn malým počtem pozorování této kategorie. I u ostatních proměnných jsou ale odhady poměrů věrohodností vysoké a jejich intervaly spolehlivosti velmi široké. Příčinou může být opět počet pozorování, a to malý počet pozorování v rámci jedné skupiny závislých pozorování a zároveň velký počet skupin. V našem případě máme totiž 170 pacientů a pro každého z nich pouze 1 až 3 měření.

Signifikance proměnných zůstala podobná jako u úplného modelu klasické logistické regrese pro tuto miRNA, tj. proměnné Grading, Resekce, Lokace meningeomu a Normované Cp jsou signifikantní, naopak proměnná Pohlaví je výrazně nesignifikantní. Pro danou hodnotu náhodného absolutního členu a při zafixování ostatních proměnných je šance rekurence meningeomu do 8 let od resekce v průměru 100krát vyšší pro pacienty s atypickým či zhoubným meningeomem (WHO grade II a III) než pro pacienty s nezhoubným meningeomem (WHO grade I). Obdobně pro pacienty, kteří podstoupili pouze částečnou resekci, je šance rekurence meningeomu 444krát vyšší než pro pacienty, kterým byl odoperován celý meningeom.

V tabulce přibyl řádek odpovídající odhadu parametru σ, tedy směrodatné

odchylky náhodného absolutního členu. Bodový odhad tohoto parametru je roven 5.01, což je poměrně vysoká hodnota. Ta nám napovídá, že mezi jednotlivými pacienty jsou velmi velké rozdíly, které mohou také přispívat ke zhoršení přesnosti odhadů. Jelikož interval spolehlivosti pro σ neobsahuje nulu, můžeme říct, že rozptyl náhodného členu je nenulový, tj. náhodný absolutní člen je signifikantní a jeho zařazení do modelu bylo správné.

Nyní přistoupíme k sestupné krokové selekci na základě AIC, abychom vybrali optimální složení proměnných v modelu. Výsledky zjednodušeného logistického modelu s náhodným absolutním členem nalezneme v Tabulce 4.9.

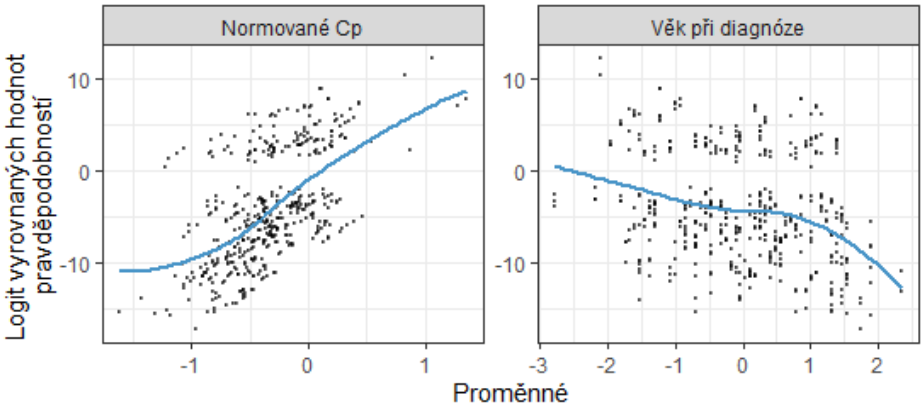
Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodn.
Absolutní člen	-10.34	(-14.45; -6.22)	-	-	< 0.0001
Věk při diagn.	-1.17	(-2.50; 0.16)	0.31	(0.08; 1.18)	0.0856
Grading - II-III	4.62	(1.55; 7.70)	101.67	(4.70; 2200.64)	0.0032
Resekce - Část.	6.05	(2.57; 9.52)	422.24	(13.05; 13665.54)	0.0007
Lokace - Conv.	4.27	(1.00; 7.54)	71.34	(2.71; 1877.57)	0.0105
Histog. - Mez.	3.51	(-0.03; 7.06)	33.60	(0.97; 1160.22)	0.0518
Normované Cp	3.17	(1.85; 4.49)	23.86	(6.38; 89.31)	< 0.0001
Parametr σ	4.97	(3.82; 6.75)	-	-	-

Tabulka 4.9: Tabulka výsledků zjednodušeného logistického modelu s náhodným absolutním členem pro *miR-331-3p*

Výsledky se odstraněním proměnné Pohlaví zpřesnily, k výrazným změnám ale nedošlo. Díky tomu, že jsme z modelu vyřadili pouze jednu proměnnou, a to Pohlaví, test podmodelu odpovídá testu významnosti této proměnné. Proto je p-hodnota testu podmodelu 0.7893 téměř totožná jako p-hodnota u proměnné Pohlaví v Tabulce 4.8. Nepatrná odlišnost je důsledkem toho, že funkce `glmer` používá k testování významnosti Waldův test, zatímco test podmodelu je test poměru věrohodností. Nulovou hypotézu o nulovosti parametru v každém případě nezamítáme a dále budeme pracovat se zjednodušeným modelem.

Pokročme tedy k diagnostice a hodnocení zjednodušeného modelu. Nejprve ověříme linearitu vztahu mezi logitem pravděpodobnosti rekurence meningeomu

a hodnotami kvantitativních proměnných Věk při diagnóze a Normované Cp (viz Obrázek 4.4).



Obrázek 4.4: Graf vztahu logitu vyrovnaných hodnot pravděpodobností a hodnot proměnných Věk při diagnóze a Normované Cp zjednodušeného logistického modelu s náhodným absolutním členem pro *miR-331-3p*

Vidíme, že ani jeden ze vztahů není na první pohled bezpochyby lineární. Zkusili jsme proto přidat do modelu druhé mocniny těchto proměnných a u proměnné Normované Cp vyšla jako signifikantní (p-hodnota 0.0287). Přidání druhé mocniny do modelu ale vedlo k celkem zásadnímu zhoršení přesnosti odhadů. Proto jsme se rozhodli pokračovat v další analýze zjednodušeného modelu, který jsme uvažovali doteď.

Dalším krokem je výpočet hodnot VIF pro ověření multikolinearity, ty nalezneme v Tabulce 4.10. Hodnoty VIF jsou v některých případech vyšší než u klasického logistického modelu. Stále se ale jedná o hodnoty menší než 2, což znamená, že multikolinearita se v datech nevyskytuje. Jako poslední spočítáme Hosmerův-Lemeshowův test dobré shody pro celkové ohodnocení modelu. Výsledná p-hodnota je rovna 0.5729, nulovou hypotézu o správnosti modelu tudíž nelze zamítnout.

Totožný postup jsme použili i pro zbylých 6 modelů. Tabulky výsledků úplných logistických modelů s náhodným absolutním členem lze najít v příloze C v Tabulkách C.1 až C.6. Výsledky zjednodušených modelů nalezneme v příloze D v Tabulkách D.1 až D.5, v jejichž popiscích uvádíme i p-hodnoty testů pod-

Proměnná	VIF
Věk při diagnóze	1.1660
Grading	1.2054
Resekce	1.2149
Lokace	1.9619
Histogeneze	1.9255
Normované Cp	1.4069

Tabulka 4.10: Hodnoty VIF pro proměnné zjednodušeného logistického modelu s náhodným absolutním členem pro *miR-331-3p*

modelu. Nejmenší p-hodnota je rovna 0.1685, a to u modelu pro *miR-1271-5p*.

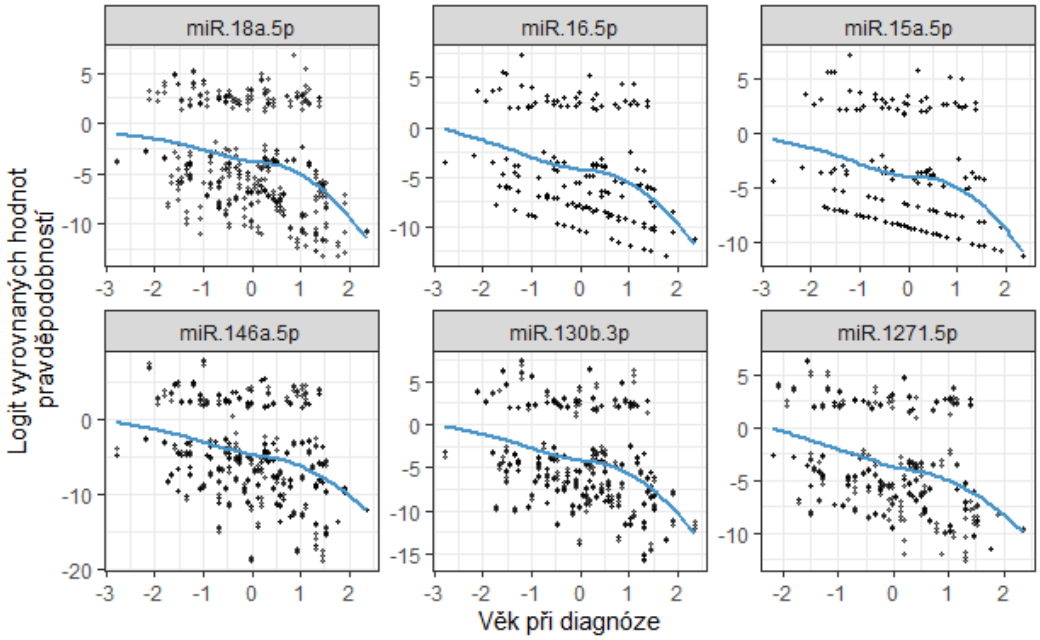
Stejně jako u *miR-331-3p*, ani u zbylých modelů nepozorujeme zásadní změny oproti výsledkům klasické logistické regrese. Resekce meningeomu je signifikantní proměnnou ve všech modelech. Pacienti, kteří podstoupili částečnou resekci meningeomu mají vyšší šanci rekurence než pacienti, kterým byl odebrán celý nádor. Stejně tak je tomu i u proměnné Grading meningeomu. Pacienti s atypickým či zhoubným nádorem mají vyšší šanci rekurence meningeomu než pacienti s nezhoubným nádorem. Proměnná Věk při diagnóze byla signifikantní ve všech modelech až na model pro *miR-18a-5p*. Naopak proměnná Pohlaví není signifikantní ani v jednom z modelů a jedná se o nejčastěji vyřazovanou proměnnou při tvorbě zjednodušených modelů. Z modelu pro *miR-130b-3p* nebyla pomocí krokové selekce vyřazena ani jedna z proměnných.

Pro nás nejzajímavější proměnná Normované Cp byla stejně jako u klasické logistické regrese signifikantní v modelech pro *miR-331-3p*, *miR-18a-5p*, *miR-146a-5p* a miR-130b-3p. Změna nastala v modelu pro *miR-1271-5p*, jelikož na rozdíl od klasického modelu, v modelu s náhodným absolutním členem vyšla proměnná Normované Cp pro tuto microRNA nesignifikantně. V obou případech jsme ale vždy pozorovali p-hodnoty velmi blízké hladině významnosti 0.05. V konečném důsledku proto ani tento výsledek není závratně odlišný od výsledku klasické logistické regrese.

Odhady směrodatné odchylky náhodného absolutního členu se u všech modelů

pohybují okolo hodnoty 5 a ani jeden z intervalů spolehlivosti neobsahuje nulu. Všechny náhodné efekty jsou tedy signifikantní.

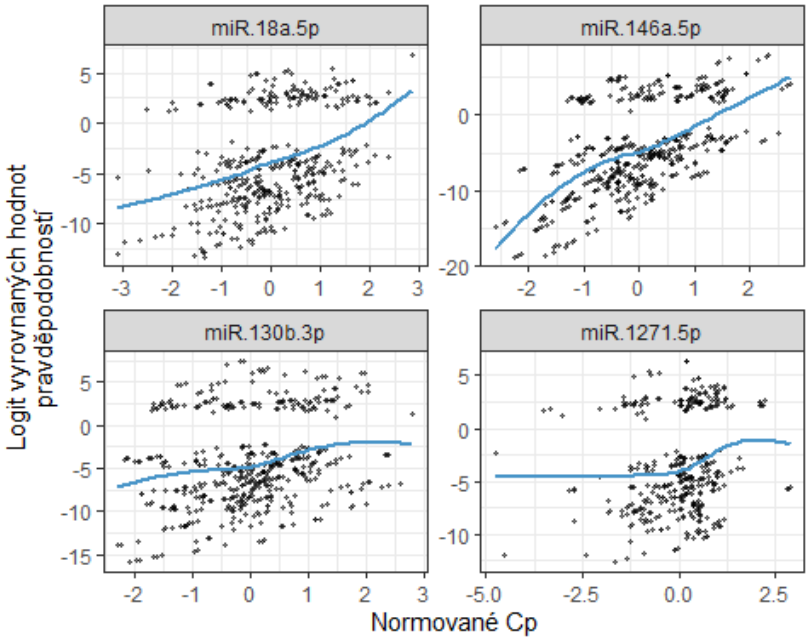
Po sestavení zjednodušených modelů opět ověříme předpoklady a zhodnotíme kvalitu modelu. Grafy vztahů mezi logitem pravděpodobností a hodnotami proměnné Věk při diagnóze vidíme na Obrázku 4.5, obdobně pro proměnnou Normované Cp na Obrázku 4.6. Opět si nemůžeme být zcela jistí linearitou všech vztahů. Druhé mocniny proměnných Věk při diagnóze a Normované Cp ale nevyšly signifikantně ani v jednom z modelů.



Obrázek 4.5: Graf vztahu logitu vyrovnaných hodnot pravděpodobností a hodnot proměnné Věk při diagnóze zjednodušených modelů s náh. absolutním členem

Spočítejme nyní hodnoty VIF pro zjednodušené modely (Tabulka 4.11). Téměř všechny hodnoty jsou menší než 2, s multikolinearitou tudíž nemáme problém.

Jako poslední krok spočítáme Hosmerův-Lemeshowův test dobré shody. V Tabulce 4.12 nalezneme p-hodnoty pro všechny zbylé modely. Nejmenší p-hodnota je rovna 0.1355, u žádného z modelů proto nezamítáme nulovou hypotézu o správnosti modelu.



Obrázek 4.6: Graf vztahu logitu vyrovnaných hodnot pravděpodobností a hodnot proměnné Normované Cp zjednodušených logistických modelů s náhodným absolutním členem

Proměnná	18a-5p	16-5p	15a-5p	146a-5p	130b-3p	1271-5p
Věk při diagn.	1.1615	1.2655	1.2409	1.1158	1.2070	1.2159
Pohlaví	-	1.1410	1.4001	-	1.0739	-
Grading	1.2411	1.1448	1.7658	1.2902	1.1437	1.2515
Resekce	1.1016	1.0699	1.6653	1.0577	1.1473	1.1190
Lokace	1.5678	1.7707	-	1.5258	2.0809	1.5956
Histogeneze	1.6434	1.8173	-	1.6931	1.6495	1.6456
Norm. Cp	1.0953	-	-	1.4422	1.3873	1.1614

Tabulka 4.11: Hodnoty VIF pro proměnné zjednodušených logistických modelů s náhodným absolutním členem (V prvním řádku tabulky jsou označení microRNA uvedeny bez předpony *miR*)

Model	p-hodnota
miR-18a-5p	0.3666
miR-16-5p	0.4400
miR-15a-5p	0.3534
miR-146a-5p	0.5033
miR-130b-3p	0.4980
miR-1271-5p	0.1355

Tabulka 4.12: Výsledné p-hodnoty Hosmerova-Lemeshowova testu dobré shody zjednodušených logistických modelů s náhodným absolutním členem

Závěr

Nejprve jsme se seznámili s teoretickým základem logistické regrese a logistické regrese s náhodným absolutním členem. Následně jsme prakticky využili nabyté znalosti a modelovali jsme závislost binární proměnné Rekurence meningeomu do 8 let od poslední resekce na řadě dalších proměnných, a to především na výsledku kvantitativní PCR metody, Normovaném Cp.

Ze zkoumaných microRNA by na základě výsledků logistické regrese jako potenciální biomarkery rekurence meningeomu mohly sloužit především *miR-331-3p* a *miR-146a-5p*. Obě zmíněné microRNA vyšly signifikantně i na základě výsledků analýzy přežití publikovaných v článku [18]. Mimo tyto dvě microRNA byla v článku prvotně signifikantní i *miR-15a-5p*, ta ale neprošla přes validační fázi studie. V našich výpočtech vyšla tato microRNA jako velmi silně nesignifikantní. Tento rozdíl může být dán použitím dvou odlišných přístupů, analýzy přežití a logistické regrese.

Z dalších faktorů byla nejčastěji signifikantním identifikátorem pacientů s vyšším rizikem rekurence meningeomu proměnná Resekce meningeomu a Grading meningeomu. Pacienti, kterým byla odoperována pouze část nádoru, mají vyšší riziko rekurence než pacienti, kterým byl odoperován celý nádor. Obdobně pacienti se zhoubným nádorem (WHO grade III) mají vyšší riziko rekurence než pacienti s nezhoubným nádorem (WHO grade I).

Výstupy klasické logistické regrese a logistické regrese s náhodným absolutním členem jsou na základě našich výpočtů velmi podobné. Zdá se tedy, že pro tato konkrétní data agregace opakovaných měření pomocí aritmetického průměru nezpůsobila výrazné změny ve výsledcích jako takových. Musíme ale podotknout,

že logistické modely s náhodným absolutním členem jsou oproti klasickým modelům značně nepřesné. Pro další praktické použití by tedy bylo nezbytné modely dále podrobně analyzovat, případně zvážit použití složitějších modelů, např. s vícero náhodnými efekty. Zmíněná nepřesnost výsledků může pramenit z velké variability mezi jednotlivými pacienty, malého množství pozorování pro každého pacienta a zároveň velkého množství pacientů.

Na pozadí logistické regrese se smíšenými efekty se skrývá velké množství složitých numerických výpočtů. My jsme v rámci praktické části narazili na problém s konvergencí. Naštěstí byl tento problém snadno vyřešen změnou optimalizační metody, kterou funkce `glmer` používá. Řada odborníků kvůli těmto úskalím ale často doporučuje pro práci se smíšenými modely upřednostňovat před softwarem R softwary lépe vybavené pro numerickou optimalizaci, jako je např. Matlab.

Práce s daty z medicínského prostředí mě velmi bavila. Doufám, že i mé výsledky alespoň z malé části přispějí k nalezení spolehlivých biomarkerů rekurence meningeomu, které následně pomohou zlepšit léčbu pacientů trpících tímto onemocněním .

Přílohy

A. Tabulky výsledků úplných logistických modelů

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-2.21	(-3.40; -1.16)	-	-	0.0001
Věk při diagnóze	-0.31	(-0.73; 0.10)	0.74	(0.48; 1.10)	0.1420
Pohlaví - Žena	-0.37	(-1.22; 0.49)	0.69	(0.30; 1.63)	0.3911
Grading - II	0.74	(-0.21; 1.67)	2.09	(0.81; 5.32)	0.1210
Grading - III	1.57	(-0.01; 3.17)	4.78	(0.99; 23.80)	0.0495
Resekce - Částečná	1.37	(0.47; 2.29)	3.93	(1.61; 9.88)	0.0029
Lokace - Convexity	1.10	(0.14; 2.15)	3.02	(1.15; 8.60)	0.0297
Histog. - Mezoderm	0.89	(-0.23; 2.04)	2.43	(0.80; 7.66)	0.1212
Normované Cp	0.48	(0.06; 0.93)	1.62	(1.06; 2.53)	0.0294

Tabulka A.1: Tabulka výsledků úplného logistického modelu pro *miR-18a-5p*

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-1.72	(-2.77; -0.78)	-	-	0.0007
Věk při diagnóze	-0.39	(-0.80; -0.01)	0.67	(0.45; 0.99)	0.0486
Pohlaví - Žena	-0.62	(-1.43; 0.19)	0.54	(0.24; 1.21)	0.1302
Grading - II	0.73	(-0.21; 1.65)	2.07	(0.81; 5.19)	0.1207
Grading - III	1.33	(-0.04; 2.73)	3.78	(0.96; 15.37)	0.0565
Resekce - Částečná	1.29	(0.44; 2.15)	3.62	(1.55; 8.55)	0.0030
Lokace - Convexity	0.77	(-0.13; 1.74)	2.17	(0.88; 5.68)	0.1022
Histog. - Mezoderm	0.66	(-0.38; 1.73)	1.94	(0.68; 5.64)	0.2138
Normované Cp	0.03	(-0.35; 0.43)	1.03	(0.70; 1.54)	0.8690

Tabulka A.2: Tabulka výsledků úplného logistického modelu pro *miR-16-5p*

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-1.63	(-2.68; -0.69)	-	-	0.0012
Věk při diagnóze	-0.38	(-0.79; 0.01)	0.68	(0.45; 1.01)	0.0581
Pohlaví - Žena	-0.66	(-1.46; 0.14)	0.52	(0.23; 1.15)	0.1027
Grading - II	0.74	(-0.21; 1.67)	2.09	(0.81; 5.30)	0.1203
Grading - III	1.28	(-0.09; 2.68)	3.60	(0.91; 14.60)	0.0655
Resekce - Částečná	1.33	(0.48; 2.20)	3.80	(1.62; 9.05)	0.0022
Lokace - Convexity	0.72	(-0.19; 1.69)	2.06	(0.83; 5.42)	0.1297
Histog. - Mezoderm	0.59	(-0.47; 1.66)	1.80	(0.63; 5.26)	0.2756
Normované Cp	-0.04	(-0.43; 0.35)	0.96	(0.65; 1.42)	0.8537

Tabulka A.3: Tabulka výsledků úplného logistického modelu pro *miR-15a-5p*

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-2.69	(-3.98; -1.57)	-	-	< 0.0001
Věk při diagnóze	-0.42	(-0.86; 0.00)	0.66	(0.42; 1.00)	0.0554
Pohlaví - Žena	-0.17	(-1.03; 0.71)	0.84	(0.36; 2.04)	0.7001
Grading - II	1.05	(0.06; 2.05)	2.86	(1.06; 7.80)	0.0369
Grading - III	2.45	(0.87; 4.11)	11.55	(2.38; 60.88)	0.0027
Resekce - Částečná	1.07	(0.16; 1.99)	2.91	(1.18; 7.28)	0.0210
Lokace - Convexity	1.24	(0.25; 2.31)	3.47	(1.29; 10.10)	0.0170
Histog. - Mezoderm	1.26	(0.14; 2.43)	3.52	(1.15; 11.41)	0.0304
Normované Cp	1.06	(0.58; 1.61)	2.90	(1.78; 5.02)	0.0001

Tabulka A.4: Tabulka výsledků úplného logistického modelu pro *miR-146a-5p*

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-2.08	(-3.28; -1.04)	-	-	0.0002
Věk při diagnóze	-0.42	(-0.84; -0.02)	0.66	(0.43; 0.98)	0.0456
Pohlaví - Žena	-0.56	(-1.40; 0.27)	0.57	(0.25; 1.31)	0.1830
Grading - II	0.87	(-0.09; 1.83)	2.40	(0.91; 6.25)	0.0725
Grading - III	1.80	(0.32; 3.35)	6.03	(1.37; 28.39)	0.0184
Resekce - Částečná	1.42	(0.52; 2.34)	4.12	(1.67; 10.40)	0.0022
Lokace - Convexity	1.13	(0.14; 2.19)	3.08	(1.15; 8.94)	0.0300
Histog. - Mezoderm	0.71	(-0.42; 1.86)	2.03	(0.65; 6.44)	0.2208
Normované Cp	0.49	(0.05; 0.94)	1.62	(1.05; 2.57)	0.0312

Tabulka A.5: Tabulka výsledků úplného logistického modelu pro *miR-130b-3p*

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-2.19	(-3.49; -1.06)	-	-	0.0004
Věk při diagnóze	-0.34	(-0.78; 0.09)	0.71	(0.46; 1.09)	0.1232
Pohlaví - Žena	-0.37	(-1.25; 0.54)	0.69	(0.29; 1.72)	0.4199
Grading - II	0.86	(-0.10; 1.82)	2.36	(0.90; 6.16)	0.0764
Grading - III	2.14	(0.30; 4.07)	8.52	(1.35; 58.76)	0.0236
Resekce - Částečná	1.18	(0.26; 2.12)	3.26	(1.30; 8.33)	0.0119
Lokace - Convexity	1.05	(0.05; 2.13)	2.85	(1.05; 8.37)	0.0463
Histog. - Mezoderm	0.99	(-0.15; 2.18)	2.70	(0.86; 8.81)	0.0913
Normované Cp	0.49	(0.01; 1.02)	1.63	(1.01; 2.78)	0.0558

Tabulka A.6: Tabulka výsledků úplného logistického modelu pro *miR-1271-5p*

B. Tabulky výsledků zjednodušených logistických modelů

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-2.46	(-3.46; -1.61)	-	-	< 0.0001
Grading - II	0.72	(-0.21; 1.64)	2.05	(0.81; 5.16)	0.1260
Grading - III	1.58	(0.07; 3.07)	4.84	(1.08; 21.57)	0.0350
Resekce - Částečná	1.44	(0.56; 2.35)	4.22	(1.75; 10.52)	0.0015
Lokace - Convexity	1.14	(0.18; 2.19)	3.13	(1.20; 8.90)	0.0242
Histog. - Mezoderm	0.91	(-0.18; 2.05)	2.49	(0.83; 7.75)	0.1046
Normované Cp	0.52	(0.12; 0.96)	1.69	(1.12; 2.62)	0.0147

Tabulka B.1: Tabulka výsledků zjednodušeného logistického modelu pro *miR-18a-5p*, p-hodnota testu podmodelu 0.2571

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-1.18	(-1.93; -0.49)	-	-	0.0013
Věk při diagnóze	-0.39	(-0.79; -0.01)	0.68	(0.45; 0.99)	0.0492
Pohlaví - Žena	-0.66	(-1.45; 0.12)	0.52	(0.24; 1.13)	0.0949
Grading - II	0.81	(-0.10; 1.72)	2.26	(0.90; 5.56)	0.0771
Grading - III	1.28	(-0.06; 2.65)	3.61	(0.94; 14.19)	0.0593
Resekce - Částečná	1.24	(0.42; 2.08)	3.46	(1.52; 7.97)	0.0031

Tabulka B.2: Tabulka výsledků zjednodušeného logistického modelu pro *miR-16-5p*, p-hodnota testu podmodelu 0.3902

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-1.14	(-1.89; -0.45)	-	-	0.0018
Věk při diagnóze	-0.37	(-0.77; 0.01)	0.69	(0.46; 1.01)	0.0595
Pohlaví - Žena	-0.70	(-1.49; 0.09)	0.50	(0.23; 1.09)	0.0817
Grading - II	0.82	(-0.11; 1.73)	2.27	(0.90; 5.64)	0.0783
Grading - III	1.24	(-0.10; 2.61)	3.47	(0.90; 13.64)	0.0680
Resekce - Částečná	1.30	(0.47; 2.14)	3.66	(1.60; 8.52)	0.0022

Tabulka B.3: Tabulka výsledků zjednodušeného logistického modelu pro *miR-15a-5p*, p-hodnota testu podmodelu 0.4441

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-2.83	(-3.91; -1.90)	-	-	< 0.0001
Věk při diagnóze	-0.41	(-0.85; 0.01)	0.66	(0.43; 1.01)	0.0600
Grading - II	1.07	(0.09; 2.07)	2.93	(1.10; 7.93)	0.0316
Grading - III	2.50	(0.94; 4.14)	12.15	(2.55; 62.87)	0.0020
Resekce - Částečná	1.07	(0.17; 1.99)	2.91	(1.18; 7.28)	0.0204
Lokace - Convexity	1.25	(0.27; 2.32)	3.51	(1.31; 10.18)	0.0158
Histog. - Mezoderm	1.28	(0.16; 2.45)	3.60	(1.18; 11.60)	0.0271
Normované Cp	1.09	(0.61; 1.62)	2.96	(1.84; 5.07)	< 0.0001

Tabulka B.4: Tabulka výsledků zjednodušeného logistického modelu pro *miR-146a-5p*, p-hodnota testu podmodelu 0.7008

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-1.75	(-2.74; -0.86)	-	-	0.0002
Věk při diagnóze	-0.41	(-0.83; -0.01)	0.66	(0.44; 0.99)	0.0488
Pohlaví - Žena	-0.61	(-1.44; 0.22)	0.54	(0.24; 1.25)	0.1468
Grading - II	0.88	(-0.07; 1.83)	2.42	(0.93; 6.25)	0.0658
Grading - III	1.78	(0.29; 3.33)	5.91	(1.34; 27.92)	0.0201
Resekce - Částečná	1.48	(0.59; 2.39)	4.37	(1.80; 10.94)	0.0012
Lokace - Convexity	0.80	(-0.03; 1.66)	2.22	(0.97; 5.27)	0.0628
Normované Cp	0.43	(0.01; 0.88)	1.54	(1.01; 2.40)	0.0476

Tabulka B.5: Tabulka výsledků zjednodušeného logistického modelu pro *miR-130b-3p*, p-hodnota testu podmodelu 0.2194

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-2.51	(-3.55; -1.61)	-	-	< 0.0001
Věk při diagnóze	-0.30	(-0.74; 0.11)	0.74	(0.48; 1.12)	0.1576
Grading - II	0.91	(-0.03; 1.86)	2.49	(0.97; 6.44)	0.0572
Grading - III	2.32	(0.52; 4.21)	10.13	(1.68; 67.07)	0.0122
Resekce - Částečná	1.21	(0.29; 2.14)	3.34	(1.34; 8.48)	0.0100
Lokace - Convexity	1.10	(0.11; 2.17)	3.01	(1.12; 8.78)	0.0342
Histog. - Mezoderm	1.06	(-0.08; 2.23)	2.87	(0.93; 9.32)	0.0704
Normované Cp	0.55	(0.08; 1.06)	1.73	(1.08; 2.89)	0.0282

Tabulka B.6: Tabulka výsledků zjednodušeného logistického modelu pro *miR-1271-5p*, p-hodnota testu podmodelu 0.4229

C. Tabulky výsledků úplných logistických modelů s náhodným absolutním členem

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-9.70	(-13.36; -6.03)	-	-	< 0.0001
Věk při diagn.	-1.13	(-2.39; 0.14)	0.32	(0.09; 1.15)	0.0806
Pohlaví - Muž	1.70	(-0.85; 4.25)	5.48	(0.43; 70.14)	0.1913
Grading - II-III	3.86	(0.83; 6.90)	47.65	(2.30; 989.03)	0.0125
Resekce - Část.	5.61	(2.15; 9.07)	272.51	(8.58; 8656.47)	0.0015
Lokace - Convex.	3.11	(0.13; 6.08)	22.31	(1.14; 437.78)	0.0409
Histog. - Mezod.	2.76	(-0.49; 6.02)	15.88	(0.61; 412.28)	0.0961
Normované Cp	1.25	(0.18; 2.33)	3.50	(1.19; 10.26)	0.0224
Parametr σ	4.98	(3.81; 6.67)	-	-	-

Tabulka C.1: Tabulka výsledků úplného logistického modelu s náhodným absolutním členem pro *miR-18a-5p*

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-10.29	(-14.56; -6.02)	-	-	< 0.0001
Věk při diagn.	-1.49	(-3.01; 0.03)	0.23	(0.05; 1.03)	0.0548
Pohlaví - Muž	2.45	(-0.25; 5.14)	11.54	(0.78; 171.27)	0.0756
Grading - II-III	4.17	(0.28; 8.05)	64.49	(1.33; 3136.60)	0.0355
Resekce - Část.	6.52	(2.42; 10.62)	679.01	(11.21; 41120.61)	0.0018
Lokace - Convex.	2.25	(-0.86; 5.36)	9.50	(0.42; 213.41)	0.1564
Histog. - Mezod.	2.56	(-0.92; 6.04)	12.90	(0.40; 419.46)	0.1499
Normované Cp	0.10	(-0.99; 1.18)	1.10	(0.37; 3.26)	0.8592
Parametr σ	5.41	(4.15; 9.67)	-	-	-

Tabulka C.2: Tabulka výsledků úplného logistického modelu s náhodným absolutním členem pro *miR-16-5p*

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-9.58	(-13.21; -5.94)	-	-	< 0.0001
Věk při diagn.	-1.34	(-2.77; 0.08)	0.26	(0.06; 1.09)	0.0650
Pohlaví - Muž	2.43	(-0.31; 5.17)	11.35	(0.73; 175.89)	0.0823
Grading - II-III	4.28	(-0.28; 8.85)	72.60	(0.75; 6992.41)	0.0660
Resekce - Část.	6.37	(2.40; 10.33)	582.96	(11.05; 30743.81)	0.0017
Lokace - Convex.	1.95	(-1.25; 5.15)	7.03	(0.29; 172.58)	0.2323
Histog. - Mezod.	2.06	(-1.26; 5.38)	7.86	(0.28; 217.78)	0.2236
Normované Cp	-0.07	(-1.19; 1.05)	0.93	(0.30; 2.87)	0.9030
Parametr σ	5.24	(4.03; 8.12)	-	-	-

Tabulka C.3: Tabulka výsledků úplného logistického modelu s náhodným absolutním členem pro *miR-15a-5p*

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-11.09	(-15.07; -7.10)	-	-	< 0.0001
Věk při diagn.	-1.53	(-2.87; -0.20)	0.22	(0.06; 0.82)	0.0241
Pohlaví - Muž	1.41	(-1.33; 4.14)	4.08	(0.27; 62.89)	0.3132
Grading - II-III	5.23	(2.21; 8.24)	186.24	(9.15; 3790.65)	0.0007
Resekce - Část.	4.43	(1.05; 7.81)	83.89	(2.86; 2461.76)	0.0102
Lokace - Convex.	4.18	(1.23; 7.13)	65.36	(3.42; 1247.97)	0.0055
Histog. - Mezod.	4.91	(1.08; 8.73)	135.11	(2.94; 6199.94)	0.0120
Normované Cp	3.50	(1.99; 5.01)	33.24	(7.35; 150.34)	< 0.0001
Parametr σ	5.14	(4.03; 6.74)	-	-	-

Tabulka C.4: Tabulka výsledků úplného logistického modelu s náhodným absolutním členem pro *miR-146a-5p*

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-10.85	(-14.90; -6.79)	-	-	< 0.0001
Věk při diagn.	-1.66	(-3.07; -0.25)	0.19	(0.05; 0.78)	0.0213
Pohlaví - Muž	2.32	(-0.24; 4.88)	10.20	(0.79; 132.00)	0.0755
Grading - II-III	4.47	(1.51; 7.42)	87.02	(4.52; 1675.64)	0.0031
Resekce - Část.	6.55	(2.49; 10.61)	700.79	(12.12; 40531.93)	0.0016
Lokace - Convex.	3.91	(0.53; 7.29)	49.96	(1.70; 1467.15)	0.0233
Histog. - Mezod.	3.04	(-0.55; 6.63)	20.98	(0.58; 759.79)	0.0965
Normované Cp	1.39	(0.12; 2.65)	4.00	(1.13; 14.14)	0.0317
Parametr σ	5.13	(4.01; 6.89)	-	-	-

Tabulka C.5: Tabulka výsledků úplného logistického modelu s náhodným absolutním členem pro *miR-130b-3p*

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-9.62	(-13.59; -5.65)	-	-	< 0.0001
Věk při diagn.	-1.43	(-2.75; -0.11)	0.24	(0.06; 0.90)	0.0341
Pohlaví - Muž	1.77	(-0.81; 4.35)	5.88	(0.45; 77.35)	0.1780
Grading - II-III	4.62	(1.21; 8.03)	101.38	(3.34; 3074.18)	0.0080
Resekce - Část.	5.08	(1.39; 8.77)	160.60	(4.00; 6451.05)	0.0070
Lokace - Convex.	3.03	(-0.03; 6.09)	20.71	(0.97; 442.43)	0.0524
Histog. - Mezod.	3.36	(-0.01; 6.74)	28.83	(0.99; 842.07)	0.0509
Normované Cp	0.97	(-0.17; 2.10)	2.63	(0.85; 8.17)	0.0944
Parametr σ	5.01	(3.80; 6.90)	-	-	-

Tabulka C.6: Tabulka výsledků úplného logistického modelu s náhodným absolutním členem pro *miR-1271-5p*

D. Tabulky výsledků zjednodušených logistických modelů s náhodným absolutním členem

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-9.41	(-12.93; -5.89)	-	-	< 0.0001
Věk při diagn.	-1.11	(-2.36; 0.14)	0.33	(0.09; 1.15)	0.0820
Grading - II-III	4.05	(1.11; 7.00)	57.65	(3.03; 1096.41)	0.0070
Resekce - Část.	5.63	(2.29; 8.98)	279.13	(9.85; 7913.75)	0.0010
Lokace - Convex.	3.45	(0.62; 6.29)	31.66	(1.85; 541.15)	0.0171
Histog. - Mezod.	2.84	(-0.24; 5.91)	17.03	(0.79; 367.19)	0.0704
Normované Cp	1.38	(0.32; 2.44)	3.98	(1.38; 11.51)	0.0107
Parametr σ	5.02	(3.84; 6.75)	-	-	-

Tabulka D.1: Tabulka výsledků zjednodušeného logistického modelu s náhodným absolutním členem pro *miR-18a-5p*, p-hodnota testu podmodelu 0.1770

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-10.33	(-14.65; -6.02)	-	-	< 0.0001
Věk při diagn.	-1.51	(-3.00; -0.02)	0.22	(0.05; 0.98)	0.0477
Pohlaví - Muž	2.51	(-0.14; 5.15)	12.26	(0.87; 172.79)	0.0634
Grading - II-III	4.11	(0.41; 7.80)	60.79	(1.51; 2447.74)	0.0294
Resekce - Část.	6.55	(2.38; 10.73)	702.03	(10.76; 45804.71)	0.0021
Lokace - Conv.	2.25	(-0.79; 5.28)	9.46	(0.45; 197.09)	0.1471
Histog. - Mez.	2.58	(-0.86; 6.02)	13.19	(0.42; 409.57)	0.1411
Parametr σ	5.42	(4.15; 10.33)	-	-	-

Tabulka D.2: Tabulka výsledků zjednodušeného logistického modelu s náhodným absolutním členem pro *miR-16-5p*, p-hodnota testu podmodelu 0.8600

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-8.66	(-11.74; -5.57)	-	-	< 0.0001
Věk při diagn.	-1.16	(-2.43; 0.12)	0.31	(0.09; 1.13)	0.0755
Pohlaví - Muž	2.17	(-0.54; 4.89)	8.80	(0.58; 132.63)	0.1163
Grading - II-III	5.29	(0.70; 9.87)	197.88	(2.02; 19429.16)	0.0239
Resekce - Část.	6.85	(2.15; 11.55)	945.19	(8.59; 104042.99)	0.0043
Parametr σ	5.52	(4.16; 9.87)	-	-	-

Tabulka D.3: Tabulka výsledků zjednodušeného logistického modelu s náhodným absolutním členem pro *miR-15a-5p*, p-hodnota testu podmodelu 0.4869

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-10.55	(-14.10; -7.00)	-	-	< 0.0001
Věk při diagn.	-1.37	(-2.62; -0.11)	0.25	(0.07; 0.89)	0.0329
Grading - II-III	5.24	(2.28; 8.20)	189.36	(9.81; 3656.68)	0.0005
Resekce - Část.	4.39	(1.20; 7.57)	80.31	(3.32; 1940.03)	0.0070
Lokace - Conv.	4.24	(1.32; 7.16)	69.46	(3.76; 1283.69)	0.0044
Histog. - Mez.	4.86	(1.03; 8.68)	128.58	(2.80; 5909.73)	0.0129
Normované Cp	3.68	(2.15; 5.21)	39.48	(8.54; 182.39)	< 0.0001
Parametr σ	5.06	(3.99; 6.50)	-	-	-

Tabulka D.4: Tabulka výsledků zjednodušeného logistického modelu s náhodným absolutním členem pro *miR-146a-5p*, p-hodnota testu podmodelu 0.3073

Proměnná	$\hat{\beta}_j$	95% CI(β_j)	$\widehat{OR}_j$	95% CI(OR_j)	p-hodnota
Absolutní člen	-9.01	(-12.40; -5.61)	-	-	< 0.0001
Věk při diagn.	-1.30	(-2.58; -0.03)	0.27	(0.08; 0.97)	0.0448
Grading - II-III	4.81	(1.58; 8.05)	123.23	(4.84; 3137.14)	0.0036
Resekce - Část.	4.90	(1.40; 8.41)	134.91	(4.06; 4481.49)	0.0061
Lokace - Convex.	3.22	(0.27; 6.17)	25.08	(1.31; 480.45)	0.0325
Histog. - Mezod.	3.30	(-0.08; 6.68)	27.17	(0.92; 799.41)	0.0556
Normované Cp	1.11	(-0.03; 2.24)	3.02	(0.97; 9.40)	0.0563
Parametr σ	4.95	(3.81; 6.67)	-	-	-

Tabulka D.5: Tabulka výsledků zjednodušeného logistického modelu s náhodným absolutním členem pro *miR-1271-5p*, p-hodnota testu podmodelu 0.1685

Literatura

[1] Biomnigene, *qPCR - RTqPCR* [online].[cit. 2021-11-24]. dostupné z: https://www.biomnigene.fr/en/our-solutions/qpcr-rt-qpcr.html

[2] Cancer.Net, *Meningioma: Types of Treatment* [online].[cit. 2021-11-20]. dostupné z: https://www.cancer.net/cancer-types/meningioma/types-treatment

[3] Demidenko, E., *Mixed models: Theory and Applications with R*, 2. vydání. John Wiley & Sons, New Jersey, 2013

[4] Hall, D. B., *Mixed-Effect Models and Longitudinal Data Analysis - Lecture Notes, Part 3* [online].[cit. 2022-01-17]. dostupné z: https://faculty.franklin.uga.edu/dhall/stat-8630

[5] Harrel, F. E., *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*, 2. vydání. Spinger, New York, 2001

[6] Hebák, P., Hustopecký, J., Pecáková, I., Plašil, M., Řezanková, H., Vlach, P., Svobodová, A., *Vícerozměrné statistické metody (3)*, 2. dopl. vydání. Informatorium, Praha, 2007

[7] Hosmer, D. W., Lemeshow, S., *Applied logistic regression*, 2. vydání. John Wiley & Sons, New York, 2000

[8] Hron, K., Kunderová, P., Vencálek, O., *Základy počtu pravděpodobnosti a metod matematické statistiky*, 3. přepracované vydání. Univerzita Palackého v Olomouci, Olomouc, 2018

[9] Jiang, J., *Linear and Generalized Linear Mixed Models and Their Applications*. Spinger, New York, 2007

[10] Jiang, J., Jia, H., Chen, H., *Maximum posterior estimation of random effects in generalized linear mixed models*. Statistica Sinica, 11(1), 97–120, 2001

[11] Johns Hopkins Medicine, *Meningioma* [online].[cit. 2021-11-19]. dostupné z: https://www.hopkinsmedicine.org/health/conditions-and-diseases/meningioma

[12] Johns Hopkins Medicine, *Meningioma grading* [online].[cit. 2021-11-20]. dostupné z: https://www.hopkinsmedicine.org/health/conditions-and-diseases/meningioma-grading

[13] McCulloch, C. E., *AN INTRODUCTION TO GENERALIZED LINEAR MIXED MODELS*. Conference on Applied Statistics in Agriculture, Kansas State University, 1996

[14] McCulloch, C. E., Searle, S. R., *Generalized, Linear and Mixed Models*, 2. vydání. John Wiley & Sons, New York, 2001

[15] Pritchard, C. C., Cheng, H. H., Tewari, M., *MicroRNA profiling: approaches and considerations*. Nature Reviews Genetics, 13, 358–369, 2012

[16] Royston, P., *Profile Likelihood for Estimation and Confidence Intervals*. The Stata Journal, 7(3), 376–387, 2007

[17] Simonoff, J. S., *Analyzing categorical data*. Springer, New York, 2003

[18] Slavik, H., Balik, V., Vrbkova, J., Rehulkova, A., Vaverka, M., Hrabalek, L., Ehrmann, J., Vidlarova, M., Gurska, S., Hajduch, M., Srovnal, J., *Identification of Meningioma Patients at High Risk of Tumor Recurrence Using MicroRNA Profiling*. Neurosurgery, 87(5), 1055–1063, 2020

[19] Social Science Computing Cooperative, *Mixed Models: Testing Significance of Effects* [online].[cit. 2021-02-06]. dostupné z: https://www.ssc.wisc.edu/sscc/pubs/MM/MM$_{T}estEffects.html$

[20] Tuerlinckx, F., Rijmen, F., Verbeke, G., De Boeck, P., *Statistical inference in generalized linear mixed models: a review*. The British journal of mathematical and statistical psychology, 59(2), 225-255, 2006

[21] Wikipedie, *Nádory mozku* [online].[cit. 2021-11-20]. dostupné z: https://cs.wikipedia.org/wiki/Nádory_mozku

[22] Wikipedie, *Real-time polymerase chain reaction* [online].[cit. 2021-11-24]. dostupné z: https://en.wikipedia.org/wiki/Real-time_polymerase_chain_reaction