UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

BAKALÁŘSKÁ PRÁCE

Statistický pohled na psychickou pohodu hráčů počítačových her

Katedra matematické analýzy a aplikací matematiky
Vedoucí práce: **doc. RNDr. Eva Fišerová, Ph.D.**
Vypracoval(a): **Lukáš Vágner**
Studijní program: B1103R004 - Aplikovaná matematika
Studijní obor: Aplikovaná matematika
Specializace: Data Science
Forma studia: prezenční
Rok odevzdání: 2024

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Lukáš Vágner

Název práce: Statistický pohled na psychickou pohodu hráčů počítačových her

Typ práce: Bakalářská práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

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

Rok obhajoby práce: 2024

Abstrakt: Bakalářská práce zkoumá vztahy mezi herními návyky a úzkostnými poruchami u hráčů počítačových her pomocí logistické regrese. v teoretické části je popsána metoda logistické regrese, včetně výpočtu koeficientů a kritérií pro nalezení vhodného modelu. Praktická část analyzuje data od téměř 13,5 tisíce respondentů z celého světa, zahrnující herní návyky, demografické údaje a psychometrické dotazníky na úzkostnou poruchu (GAD-7) a sociální fóbií (SPIN). Zjištění ukazují, že riziko úzkostných poruch se zvyšuje s časem stráveným hraním, přičemž významnou roli hraje motivace hráčů. Nejrizikovější motivací je relaxace a výhra. Ženy mají vyšší šance na úzkostné poruchy než muži. Provedená studie nabízí poznatky a hypotézy pro další výzkum.

Klíčová slova: hraní počítačových her, úzkostná porucha, sociální fóbie, logistická regrese, metoda maximální věrohodnosti, GAD-7, SPIN, SINS, SWLS

Počet stran: 91

Počet příloh: 5

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Lukáš Vágner

Title: Statistical look at psychical wellbeing of video game players

Type of thesis: Bachelor's

Department: Department of Mathematical Analysis and Application of Mathematics

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

The year of presentation: 2024

Abstract: This thesis examines the relationships between gaming habits and anxiety disorders of video games players using logistic regression. The theoretical part describes the method of logistic regression, including the calculation of coefficients and criteria for finding a suitable model. The practical part analyzes data from nearly 13,500 respondents worldwide, including gaming habits, demographic data, and psychometric questionnaires on anxiety disorder (GAD-7) and social phobia (SPIN). The findings show that the risk of anxiety disorders increases with the time spent gaming, with player motivation playing a significant role. The most risky motivations are relaxation and winning. Women have higher chances of anxiety disorders than men. The study provides insights and hypotheses for further research.

Key words: computer gaming, anxiety disorder, social phobia, logistic regression, maximum likelihood method, GAD-7, SPIN, SINS, SWLS

Number of pages: 91

Number of appendices: 5

Language: Czech

Prohlášení

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně pod vedením paní doc. RNDr. Evy Fišerové, Ph.D. a všechny použité zdroje jsem uvedl v seznamu literatury.

V Olomouci dne ....

podpis

Obsah

Úvod **10**

1 Logistická regrese **12**
1.1 Model logistické regrese a jeho vztahy k lineární regresi 12
1.2 Interpretace parametrů 15
1.2.1 Interpretace parametrů v lineární regresi 15
1.2.2 Analogie interpretace v logistické regresi 16
1.2.3 Ilustrační příklad interpretace koeficientů logistické regrese 18
1.2.4 Logistická regrese s interakcí nezávislých proměnných . 21

2 Odhad koeficientů v logistické regresi **23**
2.1 Metoda maximální věrohodnosti 23
2.2 Vlastnosti odhadů metody maximální věrohodnosti 26
2.3 Metoda maximální věrohodnosti pro výpočet koeficientů v logistické regresi 27

3 Hledání vhodného modelu **32**
3.1 Waldův test 32
3.2 Informační kritérium AIC 34
3.3 Pseudokoeficienty determinace 35

4 Hodnocení modelů binární klasifikace **37**
4.1 Klasifikační tabulka 37
4.2 ROC křivka 39

5 Praktická část **42**
5.1 Psychometrické dotazníky 42
5.2 GAD7 - generalizovaná úzkostná porucha 43
5.3 SPIN - sociálně-úzkostná porucha 44
5.4 SINS - narcismus 44
5.5 Metrika SWLS - spokojenost se životem 45
5.6 Čištění dat a popisná statistika 46
5.6.1 Rozdělení diagnóz na základě dotazníku GAD-7 46
5.6.2 Rozdělení diagnóz na základě dotazníku SPIN 48
5.6.3 Ostatní psychometrické proměnné 49
5.6.4 Demografické informace 51
5.6.5 Herní charakteristiky 56
5.7 Hledání modelu 61
5.8 Model úzkostých poruch 63

5.9 Model sociální fóbie 70

Závěr **79**

Literatura **82**

Přílohy **87**

Seznam obrázků

1 ROC křivky nejjednoduššího modelu (vlevo), modelu s pomocnou proměnnou (uprostřed) a nejkomplexnějšího modelu (vpravo) . . . 40
2 Proporční četnosti respondentů podle výskytu úzkostné poruchy diagnostikované na základě skórů GAD-7. . . . 47
3 Proporční četnosti respondentů na základě odpovědí týkajících se komplikací způsobených úzkostmi. . . . 48
4 Proporční četnosti respondentů podle výskytu sociální fóbie, diagnostikované na základě dotazníku SPIN . . . 49
5 Proporční četnosti respondentů podle výsledků SINS (vlevo) a po sdružení do nových kategorií (vpravo). . . . 50
6 Histogram respondentů podle dosaženého skóru SWLS (vlevo) a proporční četnosti po sdružení do kategorií (vpravo). . . . 51
7 Histogram respondentů dle jejich věku (vlevo) a krabicový graf (vpravo). . . . 52
8 Proporční četnosti respondentů podle jejich pohlaví. . . . 53
9 Proporční četnosti respondentů podle jejich nejvyššího dosaženého vzdělání. . . . 54
10 Proporční četnosti vzdělání respondentů po přeskupení. . . . 55
11 Proporční četnosti respondentů podle kontinentu, ze kterého pochází. . . . 56
12 Histogram (vlevo) a krabicový graf (vpravo) respondentů podle průměrného času stráveného hraním za týden. . . . 57
13 Histogram (vlevo) a krabicový graf (vpravo) respondentů podle průměrného času stráveného sledováním streamů za týden. . . 58
14 Proporční četnosti motivů ke hraní her. . . . 59
15 Proporční četnosti platforem, na kterých respondenti primárně hrají. . . . 60
16 Proporcionální četnosti typu spoluhráčů, se kterými respondent hraje nejčastěji. . . . 61
17 AIC, pseudokoeficienty determinace a AUC modelů úzkostí . . 63
18 Přesnost odhadu, senzitivita a specificita modelů úzkostí . . . 64
19 Model úzkostných poruch . . . 65
20 Koeficienty modelu úzkostných poruch a jejich vyexponovaná podoba . . . 66
21 AIC, pseudokoeficienty determinace a AUC modelů sociální fóbie . . . 70
22 Přesnost odhadu, senzitivita a specificita modelů sociální fóbie 70
23 Model sociální fóbie . . . 71

24 Model sociální fóbie 72

Poděkování

Rád bych poděkoval své vedoucí doc. RNDr. Evě Fišerové, Ph.D. za odborné rady a především trpělivost, kterou se mnou během psaní práce měla. Dále bych chtěl poděkovat svým věrným přátelům a rodině, kteří mě pocelou dobu práce i studia aktivně podporovali. Především však mé díky patří psycholožce Mgr. Heleně Maslaňákové, která mě v dobách nouze držela nad vodou a motivovala, abych své úsilí nevzdával. Těmto lidem jsem z celého srdce vděčný.

Úvod

S příchodem nového tisíciletí se prudce zvýšila dostupnost počítačů, mobilních zařízení a herních konzolí, čímž se hraní her stalo běžným koníčkem mladých generací. Stejně tak se v posledních desítkách let zásadně vyvinuly strategie herních vývojářů, aby hry hráče co nejvíc bavily. Hry nyní obsahují čím dál více úkolů ke splnění, odměn ke získání, výzev k překonání a všemožných dovedností, ve kterých se hráč může vylepšovat. Hráči tak tráví často několik hodin denně ve virtuálních světech, což vzbudilo zájem odborníků, kteří zkoumají, jaké vlivy má hraní her na psychiku a zda nepředstavuje rizika pro duševní zdraví.

Ke konci prvního desetiletí se začaly objevovat první studie se špatnými zprávami. Bylo například zjištěno, že je hraní her spojeno s psychofyziologickými změnami, které mohou dlouhodobě ovlivňovat mechanismy v těle a mozku zodpovědné za pocit potěšení a motivace podobným způsobem jako užívání návykových látek. [31] v roce 2013 bylo v DSM-5 (diagnostický a statistický manuál duševních poruch)[3] navrženo další zkoumání možné závislosti na počítačových hrách a v roce 2018 bylo nasbíráno dostatek důkazů pro oficiální zařazení této závislosti do MKN-11 (Mezinárodní klasifikace nemocí)[32], což dále prohloubilo zájem výzkumníků o tuto problematiku.

Mimo jiné byl výzkumníky zjištěn i vztah mezi excesivním hraním her a úzkostmi, depresemi a nízkou životní spokojeností[15], dále byly u hráčů zjištěny nižší sociální dovednosti[33], nebo že lidé se sociální fóbií často splňují kritéria pro závislost na hraní.[7]

Výzkum však obvykle probíhal na malých vzorcích, což vedlo ke vzniku datové sady, která je základním stavebním kamenem této práce. Výzkumníkům z mnichovské univerzity [23] se podařilo pomocí dotazníku vyvěšeného

na internetu nasbírat data od téměř 13,5 tisíce respondentů z celého světa. Kromě otázek na herní návyky a demografické informace, byly respondentům předloženy psychometrické inventáře na úzkostnou poruchu (GAD-7) a sociální fóbii (SPIN).

Tato práce si klade za cíl prozkoumat spojitostí mezi herními návyky a úzkostnými poruchami, přičemž speciální pozornost věnujeme sociální fóbii, která pod úzkostné poruchy spadá. Analýza bude probíhat pomocí modelů logistické regrese, kterým se budeme věnovat v teoretické části, kde si představíme základní koncepty logistické regrese, výpočet koeficientů pomocí metody maximální věrohodnosti, způsob interpretace koeficientů a některá kritéria pro porovnávání modelů. Následně tyto znalosti aplikujeme na dříve zmíněnou datovou sadu v praktické části, kde budou výsledky interpretovány a představeny možné hypotézy vyplývající z těchto výsledků.

Práce by mohla být především přínosná pro ty, kteří se snaží proplout do hlubin logistické regrese, či pro psychology, kteří se zajímají o problematiku hraní her.

1 Logistická regrese

Tato kapitola představí základní teoretické principy logistické regrese, konstrukci modelu a související pojmy. Následně se zaměříme na interpretaci koeficientů, přičemž začneme u lineární regrese, kde jsou principy podobné těm v logistické regresi, ale snazší na pochopení. Poté analogicky přejdeme k logistické regresi. Pro intuitivnější pochopení si ukážeme interpretaci na konkrétních příkladech, kde se také budeme věnovat interakcím závislých proměnných. Přitom budeme vycházet z úrovně vysokoškolských znalostí pravděpodobnosti a matematické statistiky, například [10].

1.1 Model logistické regrese a jeho vztahy k lineární regresi

Logistickou regresi můžeme chápat jako jakousi modifikaci lineární regrese. Lineární regrese modeluje střední hodnotu spojité náhodné veličiny Y pomocí skupiny nezávislých proměnných $X_1, X_2, \ldots, X_m$. Tudíž pomocí hodnot jiných veličin se snažíme odhadnout hodnotu Y. Model vypadá takto:

$$E(Y|X_1, X_2, \ldots, X_m) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m. \tag{1.1}$$

Pro i-té pozorování platí:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_m X_{im} + \varepsilon_i, \quad i = 1, \ldots, n. \tag{1.2}$$

Cílem je odhadnout hodnoty koeficientů $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_m)^T$, které nám říkají, jak proměnné ovlivňují závislou proměnnou za platnosti předpokladů:

$$E(\varepsilon_i) = 0, \quad var(\varepsilon_i) = \sigma^2, \quad cov(\varepsilon_i, \varepsilon_j) = 0, \quad \forall i \neq j, \quad i, j = 1, \ldots, n.$$

Problém nastává v okamžiku, kdy je proměnná Y kategoriální. v případě binární proměnné uvažujeme, že je z alternativního rozdělení, tedy $Y \approx Alt(\pi)$. Tato proměnná nabývá dvou hodnot podle toho, zda daný jev nastane (značíme 1) či nenastane (značíme 0). Pravděpodobnost nastání jevu je pak vyjádřena parametrem π:

$$P(Y = 1) = \pi \quad a \quad P(Y = 0) = 1 - \pi. \tag{1.3}$$

Logistická regrese má za cíl tuto pravděpodobnost π predikovat. Kdybychom odhad prováděli pomocí lineární regrese, vypadal by nějak takto:

$$\pi = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m.$$

Model by však jistě byl chybný, jelikož pravděpodobnost nastání jevu 1, která se vyskytuje na levé straně, se z definice pohybuje na intervalu $\langle 0, 1 \rangle$, zatímco hodnoty modelu na straně pravé se mohou pohybovat na celém spektru $(-\infty, \infty)$.

Prvním krokem k řešení tohoto problému je nahradit pravděpodobnost jevu šancí. Šance je definována jako poměr pravděpodobnosti, že jev nastane a pravděpodobnosti že nenastane. Formálně:

$$odds(P(Y = 1)) = \frac{P(Y = 1)}{P(Y = 0)} = \frac{\pi}{1 - \pi}. \tag{1.4}$$

Jestliže budeme v modelu odhadovat šanci místo pravděpodobnosti, model bude vypadat takto:

$$\frac{\pi}{1 - \pi} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m. \tag{1.5}$$

Šance na levé straně se pohybuje na oboru hodnot $\langle 0, \infty)$. Za předpokladu, že

pravděpodobnost úspěchu nebude nulová, můžeme na šanci použít přirozený logaritmus, jelikož je jeho definičním oborem právě interval $(0, \infty)$:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1 + \ldots + \beta_m X_m. \tag{1.6}$$

Takto jsme došli ke tvaru, se kterým už můžeme být spokojeni, jelikož obor hodnot logaritmu šance je právě $(-\infty, \infty)$ stejně jako obor hodnot na pravé straně rovnice. Zároveň se jedná o model logistické regrese. Funkce nalevo se jmenuje logitová, po které je logistická regrese pojmenována:

$$logit(\pi) = \ln\left(\frac{\pi}{1-\pi}\right). \tag{1.7}$$

Vyjádření pravděpodobnosti dosáhneme odlogaritmováním rovnice (1.6), tudíž použitím funkce e^x a sérií následujících úprav. Pro zjednodušení uvažujme lineární kombinaci v čitateli, jako skalární součin $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_m)^T$ a $\boldsymbol{X} = (1, X_1, \ldots, X_m)^T$:

$$\begin{aligned}
\frac{\pi}{1-\pi} &= e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_m X_m} = e^{\boldsymbol{\beta}^T \boldsymbol{x}} \\
\pi &= (1-\pi) e^{\boldsymbol{\beta}^T \boldsymbol{x}} \\
\pi + \pi e^{\boldsymbol{\beta}^T \boldsymbol{x}} &= e^{\boldsymbol{\beta}^T \boldsymbol{x}} \\
\pi &= \frac{e^{\boldsymbol{\beta}^T \boldsymbol{x}}}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}}}.
\end{aligned}$$

Pravděpodobnost úspěchu je tedy:

$$P(Y=1) = \pi = \frac{e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_m X_m}}{1 + e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_m X_m}}. \tag{1.8}$$

Pravděpodobnost neúspěchu lze vypočítat analogicky:

$$P(Y = 0) = 1 - \pi = \frac{1}{1 + e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_m X_m}}. \tag{1.9}$$

Jelikož bude výsledná pravděpodobnost záviset na realizacích $\boldsymbol{x}$ náhodného vektoru $\boldsymbol{X}$, budeme uvažovat podmíněnou pravděpodobnost:

$$P(Y = 1|\boldsymbol{X} = \boldsymbol{x}) = \frac{e^{\boldsymbol{\beta}^T \boldsymbol{x}}}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}}} \quad a \quad P(Y = 0|\boldsymbol{X} = \boldsymbol{x}) = \frac{1}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}}}. \tag{1.10}$$

1.2 Interpretace parametrů

1.2.1 Interpretace parametrů v lineární regresi

Pro jednoduchost si připomeňme, jak se parametry interpretují u lineární regrese. Máme-li závislou proměnnou Y a nezávislé proměnné X_1 a X_2, kde X_1 je kategoriální o dvou kategoriích a X_2 kvantitativní, model bude ve tvaru:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i.$$

Koeficient β_0 by v tomto případě říkal, jakou bude mít Y střední hodnotu, jestliže kvantitativní proměnná X_2 bude 0 a X_1 bude z první skupiny, kterou označíme jako 0. Koeficient β_1 říká, o kolik se nám hodnota Y sníží/zvýší, když X_1 bude ze druhé skupiny, kterou označíme jako 1. a nakonec β_2 ukazuje, o kolik se Y změní, když se kvantitativní proměnná zvýší o jednotku.

Například řekněme, že proměnná Y představuje hmotnost člověka v kg. Kategoriální nezávislá proměnná X_1 bude představovat pohlaví (0 žena, 1 muž)

a kvantitativní proměnná X_2 výšku v cm:

$$\hat{Y}_i = 53 + 10X_1 + 0{,}5X_2.$$

Kdyby se hypoteticky našla žena, která měří 0 cm, vážila by 53 kg. Koeficient u kategoriální proměnné vysvětluje, jak se průměrná hmotnost změní, když sledovaná osoba není žena, ale muž, a to by představovalo zvýšení hmotnosti o 10 kg. Dále s každým centimetrem do výšky se průměrná hmotnost zvýší o 0,5 kg.

1.2.2 Analogie interpretace v logistické regresi

U logistické regrese se budeme pokoušet o něco podobného. Interpretace bude však trochu složitější, jelikož nesledujeme přímo změnu na závislé proměnné, ale to, jak se bude měnit poměr šancí (OR - odds ratio), který je pro spojité proměnné formálně definován takto:

$$OR(X_i) = \frac{odds(P(Y = 1|X_1 = x_1, \ldots, X_i = x_i + 1, \ldots, X_m = x_m))}{odds(P(Y = 1|X_1 = x_1, \ldots, X_i = x_i, \ldots, X_m = x_m))} \quad (1.11)$$

a pro kategoriální:

$$OR(X_i) = \frac{odds(P(Y = 1|X_1 = x_1, \ldots, X_i = x_{i,j}, \ldots, X_m = x_m))}{odds(P(Y = 1|X_1 = x_1, \ldots, X_i = x_{i,0}, \ldots, X_m = x_m))}, \quad (1.12)$$

kde j je označení pro kategorii i-té proměnné, kterou budeme porovnávat s kategorií 0, která je součástí interceptu. Interpretace poměru šancí si podrobně rozebereme v dalších odstavcích.

V případě spojité proměnné X_1:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1,$$

budeme pozorovat, co se bude dít, když X_1 zvětšíme o jedničku:

$$\ln\left(\frac{\pi^*}{1-\pi^*}\right) = \beta_0 + \beta_1(X_1 + 1),$$

a tyto modely od sebe odečteme:

$$\ln\left(\frac{\pi^*}{1-\pi^*}\right) - \ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1(X_1 + 1) - (\beta_0 + \beta_1 X_1) = \beta_1.$$

Výsledkem je koeficient β_1. Když vztah nalevo upravíme pomocí vlastností logaritmu:

$$\ln\left(\frac{\pi^*}{1-\pi^*}\right) - \ln\left(\frac{\pi}{1-\pi}\right) = \ln\left(\frac{\frac{\pi^*}{1-\pi^*}}{\frac{\pi}{1-\pi}}\right) = \beta_1,$$

získáme v logaritmu již dříve zmíněný poměr šancí (1.11). Dosadíme-li vztah (1.10), dostaneme:

$$OR(X_i) = e^{\beta_i}.$$

Parametry $\boldsymbol{\beta}$ interpretujeme ve vyexponovaném tvaru a vyjadřují nám, kolikrát se šance nastoupení jevu změní s každým navýšením závislé proměnné o jednotku.

V případě kategoriálních proměnných bude koeficient β_0 (intercept) reprezentovat výchozí kategorii, se kterou budeme ostatní srovnávat:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0. \tag{1.13}$$

Pro každou z k kategorií se vytvoří nová proměnná, jejíž koeficient předsta-

vuje změnu na interceptu. Jestliže pozorování bude náležet do j-té kategorie, kde $j = 1, \ldots, k$ (jelikož nultá kategorie je v interceptu), potom $X_j = 1$ a všechny ostatní $X_i = 0$, kde $i = 1, \ldots, j-1, j+1, \ldots, k$. Předpokládejme, že pozorování náleží do kategorie 2:

$$\ln\left(\frac{\pi*}{1-\pi*}\right) = \beta_0 + \beta_2 X_2 = \beta_0 + \beta_2. \tag{1.14}$$

Po odečtení modelu (1.13) od (1.14) a následným uplatněním vlastností logaritmu opět získáme:

$$\ln\left(\frac{\frac{\pi^*}{1-\pi^*}}{\frac{\pi}{1-\pi}}\right) = \beta_0 + \beta_2 - \beta_0 = \beta_2.$$

Tedy:

$$OR(X_i) = e^{\beta_i},$$

nám bude říkat, kolikrát se změní poměr šancí (1.12), když budeme uvažovat i-tou kategorii místo výchozí.

1.2.3 Ilustrační příklad interpretace koeficientů logistické regrese

Ukažme si, jak můžeme koeficienty interpretovat na jednoduchém příkladu. Uvažujme opět model, kde máme mezi nezávislými proměnnými jednu kategoriální proměnnou X_1, která zahrnuje dvě kategorie, a jednu kvantitativní proměnnou X_2. Obecný model by vypadal takto:

$$\ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i.$$

Odhadujeme pravděpodobnost, že se studentům podaří dostat na vysokou školu, budou-li při přijímacím řízení psát test, ve kterém je bodová hranice, kterou když student překročí, je přijat. Kategoriální proměnná X_1 bude

představovat pohlaví (0 zastupuje ženy a 1 muže) a spojitá proměnná X_2 počet přípravných testů, které si student před přijímacím řízením vypracoval. Řekněme, že model vyšel následovně:

$$\ln\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = -0{,}6 - 0{,}56X_1 + 1{,}13X_2.$$

Samotné koeficienty nám toho moc neříkají. z vlastností exponenciální funkce, ke které přejdeme později, víme, že je-li koeficient záporný můžeme očekávat pokles na pravděpodobnosti. Když je kladný, očekáváme růst.

První koeficient β_0, který je v tomto příkladu -0,6, funguje jako intercept (odraz), což je základní úroveň, ze které model vychází. Zároveň se v interceptu schovává výchozí skupina z kategoriálních proměnných, se kterou budeme ostatní kategorie srovnávat, tedy v příkladu jsou to ženy. Jelikož je koeficient záporný, tak počáteční šance úspěchu bude nižší než neúspěchu. Koeficient u kategoriální proměnné β_1 je také záporný, což signalizuje, že šance mužů bude nižší než žen. Poslední koeficient u kvantitativní proměnné β_2 je kladný, tudíž se šance na přijetí s každým absolvovaným přípravným testem zvýší.

Pro hlubší porozumění koeficientům modelu je potřeba provést odlogaritmování, čímž přejdeme k samotnému poměru šancí. Tímto způsobem lze získat přímočařejší představu o tom, jak se šance na úspěch mění v závislosti na hodnotách nezávislých proměnných:

$$\frac{\hat{\pi}}{1-\hat{\pi}} = e^{-0{,}6-0{,}56X_1+1{,}13X_2} = e^{-0{,}6} \cdot e^{-0{,}56X_1} \cdot e^{1{,}13X_2} = 0{,}55 \cdot 0{,}57^{X_1} \cdot 3{,}09^{X_2}.$$

Stěžejním bodem u poměru šancí bude 1. Když je poměr šancí větší jak 1, očekáváme vyšší pravděpodobnost úspěchu než neúspěchu. Pravděpodobnost

úspěchu se tedy bude pohybovat nad 50 %. v opačném případě, tedy při hodnotě menší jak 1 budeme očekávat větší pravděpodobnost neúspěchu.

Poměr šancí pro intercept je v příkladu 0,55. Což znamená, že pravděpodobnost úspěchu je o 45 % nižší, než neúspěchu. Hodnotu pravděpodobnosti úspěchu vypočítáme takto:

$$P(Y = 1|X_1 = 0, X_2 = 0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}} = \frac{0{,}55}{1 + 0{,}55} = 0{,}355.$$

Uvážíme-li kontext přikladu, zjistíme, že žena, která si nevyplní jediný přípravný test, bude mít pravděpodobnost přijetí na vysokou školu 35,5 %. Její naděje na přijetí je tedy nízká.

V případě, že nebudeme uvažovat ženu, ale muže (tedy $X_1 = 1$), výpočet poměru šancí bude vypadat následovně:

$$e^{\beta_0 + \beta_1 \cdot 1} = e^{\beta_0} \cdot e^{\beta_1} = 55 \cdot 0{,}57 = 0{,}31.$$

Tento výsledek znamená, že se poměr šancí na úspěch v případě mužů ještě více sníží, konkrétně o 69 % ve srovnání se ženou. Podívejme se, jak bude vypadat pravděpodobnost na úspěch v případě mužů:

$$P(Y = 1|X_1 = 1, X_2 = 0) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}} = \frac{0{,}31}{1 + 0{,}31} = 0{,}236.$$

Muži se na vysokou školu dostanou s pravděpodobností 23,6 %, což je skutečně méně než v případě žen.

v tomto příkladu jsou skupiny jen dvě, ale pokud by kategorií bylo více, byl by pro každou další vypočítán příslušný koeficient β_i. Například, pokud bychom měli pět skupin, potřebovali bychom koeficient interceptu, který by zahrnoval první kategorii, a další čtyři koeficienty pro zbývající skupiny.

Poslední koeficient, který nám zbývá interpretovat je β_2 u spojité pro-

měnné. Poměr šancí vyšel 3,09, a to znamená, že s každým vyplněným přípravným testem se nám šance na úspěch více než trojnásobně zvýší. Při jednom vyplněném testu u ženy je poměr šancí $0{,}55 \cdot 3{,}09^1 = 1{,}7$ a u muže $0{,}31 \cdot 3{,}09^1 = 0{,}96$. u ženy přesáhl 1, takže se spíše na vysokou školu dostane. Zatímco muž 1 nepřesáhl, stále je tedy u něj pravděpodobnější neúspěch. Ověřme si to pomocí pravděpodobnosti:

$$P(Y = 1|X_1 = 0, X_2 = 1) = \frac{e^{\beta_0+\beta_2\cdot 1}}{1 + e^{\beta_0+\beta_2\cdot 1}} = \frac{1{,}7}{1 + 1{,}7} = 0{,}63$$

$$P(Y = 1|X_1 = 1, X_2 = 1) = \frac{e^{\beta_0+\beta_1\cdot 1+\beta_2\cdot 1}}{1 + e^{\beta_0+\beta_1\cdot 1+\beta_2\cdot 1}} = \frac{0{,}96}{1 + 0{,}96} = 0{,}49.$$

Nutno podotknout, že je tento příklad smyšlený, ale po takovéto analýze bychom zjistili, že ženám stačí, aby si vypracovala jeden přípravný test a už by měla nadpoloviční pravděpodobnost přijetí, zatímco muži by potřebovali vyplnit testy alespoň dva.

1.2.4 Logistická regrese s interakcí nezávislých proměnných

Můžeme se setkat se situací, kdy nám efekt samotných nezávislých proměnných nebude stačit a některé proměnné spolu interagují, což znamená, že když působí na závislou proměnnou současně, tak se jejich efekt může lišit.

Vysvětleme si to na předešlém příkladu, kde se zkoumal vliv pohlaví a počtu vyplněných přípravných testů na úspěšnost u přijímaček. Původní příklad bez interakce s odlogaritmovanými koeficienty vypadal takto:

$$\frac{\hat{\pi}}{1 - \hat{\pi}} = e^{-0{,}6-0{,}56X_1+1{,}13X_2} = 0{,}55 \cdot 0{,}57^{X_1} \cdot 3{,}09^{X_2}.$$

Z toho plynulo, že ženy mají obecně větší šanci na přijetí než muži, jelikož jejich počáteční pravděpodobnost byla vyšší. Předpokládali jsme, že se

všichni učí stejně, tudíž s každým vypočítaným testem se pravděpodobnost zvětšila 3krát. Co když ale uvážíme interakci mezi přípravou a pohlavím?

$$\frac{\hat{\pi}}{1-\hat{\pi}} = e^{-0{,}6-0{,}56X_1+0{,}9X_2+1{,}52X_2X_1} = 0{,}55 \cdot 0{,}57^{X_1} \cdot 2{,}46^{X_2} \cdot 4{,}48^{X_2X_1}.$$

Zde vidíme, že když interakci uvažovat budeme, dozvíme se, že se muži učí rychleji. Zatímco u žen se poměr šancí snížil a s každým vyplněným testem šance na přijetí vzroste 2,5krát, u můžu se tato šance zvýší 4,5krát. V původním příkladu byla pravděpodobnost přjetí ženy, když si vypracovala jeden test 63 % a muže 49 %. Co se stane po novém zjištění?

$$P(Y=1|X_1=0, X_2=1) = \frac{e^{\beta_0+\beta_2\cdot 1}}{1+e^{\beta_0+\beta_2\cdot 1}} = \frac{1{,}35}{1+1{,}35} = 0{,}57$$

$$P(Y=1|X_1=1, X_2=1) = \frac{e^{\beta_0+\beta_1\cdot+\beta_3\cdot 1}}{1+e^{\beta_0+\beta_1+\beta_3\cdot 1}} = \frac{1{,}43}{1+1{,}43} = 0{,}59.$$

Ačkoliv oba budou spíše přijati, muži mají pravděpodobnost přijetí najednou o 2 % vyšší. Interpretace se nám tedy úplně změnila, a proto se v některých případech vyplatí interakce uvažovat, vyjdou-li koeficienty statisticky významné. Dojdeme tak ke komplexnějšímu zjištění.

V tomto příkladě jsme použili interakci kategoriální proměnné a kvantitativní. Můžeme však dát do interakce proměnné v jakékoliv kombinaci. Stejně tak v této práci převážně budeme používat interakci dvojnou. Je však možné dát do interakce i více proměnných, jako například věk, pohlaví a počet testů.

2 Odhad koeficientů v logistické regresi

Po prozkoumání konstrukce modelu a interpretace jeho koeficientů se nyní posuneme k základu, na kterém stojí výpočet těchto koeficientů. u logistické regrese je pro tyto účely klíčová metoda maximální věrohodnosti. v následujících kapitolách budeme čerpat ze zdrojů [1, 2, 29, 34]

2.1 Metoda maximální věrohodnosti

V mezinárodní literatuře se metoda maximální věrohodnosti často označuje zkratkou MLE, což v angličtině znamená „maximum-likelihood estimation“. Tento termín přímo odkazuje na práci s funkcí věrohodnosti (likelihood function). Úlohou je tuto funkci maximalizovat. Máme-li náhodný výběr n nezávislých náhodných veličin $Y_1, Y_2, \ldots, Y_n$, které pochází ze stejného rozdělení s parametry $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, kde $\boldsymbol{\Theta}$ je parametrický prostor, potom budeme hledat takové odhady $\hat{\boldsymbol{\theta}}$, které maximalizují pravděpodobnost, že daná datová sada pochází z určitého rozdělení. Díky tomu se odhady parametrů co nejlépe přizpůsobí pozorovaným datům, která jsou základem modelování.

V případě, že $Y_1, Y_2, \ldots, Y_n$ pochází ze spojitého rozdělení, tato rozdělení budou charakterizována hustotou pravděpodobnosti $f(y, \boldsymbol{\theta})$. v případě diskrétního rozdělení to bude pravděpodobnostní funkce $p(y, \boldsymbol{\theta})$, kde $\boldsymbol{\theta}$ označuje vektor neznámých parametrů. Jelikož se v této práci věnujeme logistické regresi, zaměříme se na případ diskrétního rozdělení.

Sdružená pravděpodobnostní funkce pro n realizací náhodné veličiny Y se v takovém případě vyjadřuje jako součin jednotlivých pravděpodobností, neboť náhodné veličiny jsou nezávislé a mají stejné rozdělení. Tvar této funkce bude:

$$p(y_1, \ldots, y_n | \boldsymbol{\theta}) = p(\boldsymbol{y}, \boldsymbol{\theta}) = \prod_{i=1}^{n} p(y_i, \boldsymbol{\theta}). \tag{2.1}$$

V kontextu věrohodnostní funkce $\boldsymbol{\theta}$ představuje parametry modelu, zatímco $\boldsymbol{y}$ jsou proměnné, tedy hodnoty náhodných veličin. Klíčovým rozlišením mezi věrohodnostní funkcí a sdruženou pravděpodobnostní funkcí je právě v tom, jak jsou tyto role rozděleny. v případě věrohodnostní funkce dochází k záměně, kdy $\boldsymbol{\theta}$ se stává proměnnými, které se snažíme optimalizovat, a $\boldsymbol{y}$, tedy pozorované hodnoty, považuje za parametry.

Tato změna perspektivy umožňuje přistupovat k $\boldsymbol{\theta}$ jako k hodnotám, které hledáme, aby se maximalizovala pravděpodobnost (nebo přesněji věrohodnost) pozorovaných dat $\boldsymbol{y}$ při daných $\boldsymbol{\theta}$. Věrohodnostní funkce tedy vyjadřuje, jak „věrohodné“ jsou různé hodnoty parametrů $\boldsymbol{\theta}$ vzhledem k pozorovaným datům $\boldsymbol{y}$, a naším cílem je najít takové $\hat{\boldsymbol{\theta}}$, pro které je tato věrohodnost maximální.

$$L(\boldsymbol{\theta}|y_1, \ldots, y_n) = \prod_{i=1}^{n} p(y_i, \boldsymbol{\theta}). \tag{2.2}$$

Výsledným odhadem je tedy $\hat{\boldsymbol{\theta}}$, která splňuje:

$$\hat{\boldsymbol{\theta}}_{MLE} = \arg\max_{\theta \in \Theta} L(\boldsymbol{\theta}|y_1, \ldots, y_n). \tag{2.3}$$

To znamená, že jsme identifikovali odhad, jehož věrohodnost je vyšší, než věrohodnost jakéhokoliv jiného odhadu z parametrického prostoru Θ.

Často se ukazuje jako výhodnější maximalizovat logaritmus věrohodnostní funkce $\ln L(\boldsymbol{\theta})$. Tento přístup převádí původní součin pravděpodobností na součet jejich logaritmů, což je výpočetně efektivnější. Navíc vzhledem k vlastnostem logaritmu výsledek maximalizace věrohodnosti zůstává nezměněn:

$$\ln(L(\boldsymbol{\theta}|y_1, \ldots, y_n)) = \ln(\prod_{i=1}^{n} p(y_i, \boldsymbol{\theta})) = \sum_{i=1}^{n} \ln(p(y_i, \boldsymbol{\theta})).$$

Tento přístup zjednodušuje výpočty a zároveň uchovává základní princip maximalizace věrohodnosti, což je klíčové pro nalezení nejvhodnějších parametrů $\boldsymbol{\theta}$ modelu.

Abychom našli maximum věrohodnostní funkce, budeme hledat parciální derivace logaritmické věrohodnostní funkce $\ln L(\boldsymbol{\theta}|\boldsymbol{y})$ podle každé z m proměnných ve vektoru parametrů $\boldsymbol{\theta}$, přičemž předpokládáme, že alespoň druhé derivace existují. Tyto parciální derivace nastavíme rovny nule. Tím vznikne systém rovnic, kterým se říká věrohodnostní rovnice:

$$\frac{\partial \ln L(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \theta_i} = 0 \quad i \ = 1, \ldots, m.$$

Za předpokladu nalezení řešení systému věrohodnostních rovnic je dalším krokem ověření, zda toto řešení skutečně maximalizuje logaritmickou věrohodnostní funkci. Pro tuto úlohu se používá Hessova matice druhých parciálních derivací, která je definována jako:

$$\boldsymbol{H}(\hat{\boldsymbol{\theta}}) = \left(\frac{\partial^2 \ln L(\boldsymbol{\theta}|\boldsymbol{y})}{\partial \theta_i \partial \theta_j} \right)_{i,j=1}^{m}.$$

V případě, že nalezené řešení skutečně odpovídá maximu, je nezbytné, aby Hessova matice byla negativně definitní. Budeme také vycházet z obecně platného vztahu: je-li matice $\boldsymbol{A}$ pozitivně (semi)definitní, pak matice $-\boldsymbol{A}$ je negativně (semi)definitní.

Pozitivní definitnost matice znamená, že pro všechny nenulové vektory $\boldsymbol{x}$ platí, že $\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} > 0$. v případě, že chceme, aby Hessova matice byla negativně definitní, vyžadujeme, aby pro všechny nenulové vektory $\boldsymbol{x}$ platilo, že $\boldsymbol{x}^T \boldsymbol{H} \boldsymbol{x} < 0$, což indikuje, že jsme našli maximum funkce.

Tímto způsobem lze potvrdit, že nalezené parametry $\hat{\boldsymbol{\theta}}$ nejen vyhovují podmínkám věrohodnostních rovnic, ale také odpovídají maximu logarit-

mické věrohodnostní funkce, což je klíčové pro validitu modelu logistické regrese.

2.2 Vlastnosti odhadů metody maximální věrohodnosti

Pro úplné pochopení kvality odhadů získaných metodou maximální věrohodnosti (MLE) je nezbytné se podrobně seznámit s jejich klíčovými vlastnostmi. Tyto vlastnosti odhalují, proč je MLE tak oblíbená a efektivní v praxi. Klíčovým předpokladem pro aplikovatelnost uvedených vlastností je existence dostatečně velkého množství pozorování, obecně se doporučuje více než 30.

- **Princip invariance**: Tento princip naznačuje, že pokud je $\hat{\boldsymbol{\theta}}$ maximálně věrohodný odhad parametru $\boldsymbol{\theta}$, pak transformace tohoto odhadu $u(\hat{\boldsymbol{\theta}})$ je maximálně věrohodným odhadem transformace parametru $u(\boldsymbol{\theta})$. To znamená, že transformace odhadů zachovává jejich věrohodnostní kvalitu, což umožňuje flexibilitu v interpretaci a použití odhadů bez nutnosti dalších složitých výpočtů.

- **Konzistence**: Konzistentní odhad znamená, že s rostoucím počtem pozorování odhad $\hat{\boldsymbol{\theta}}$ konverguje podle pravděpodobnosti ke skutečné hodnotě $\boldsymbol{\theta}$. Tato vlastnost zajistí, že při dostatečném množství dat bude odhad blízký pravdivé hodnotě parametru.

  Platí tedy:

$$\lim_{n\to\infty} P(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 < \epsilon) = 1 \quad \forall \epsilon > 0, \forall \boldsymbol{\theta} \in \Theta.$$

- **Asymptotická eficience**: Mezi konzistentními odhady má odhad $\hat{\boldsymbol{\theta}}$ získaný metodou MLE asymptoticky nejnižší možný rozptyl. To zna-

mená, že v limitě velkého počtu pozorování poskytuje MLE nejpřesnější odhady.

- **Asymptotická normalita**: Při dostatečném počtu pozorování se rozdělení odhadů $\hat{\boldsymbol{\theta}}$ přibližuje normálnímu rozdělení. Tato vlastnost umožňuje použití standardních statistických postupů pro odhad intervalů spolehlivosti a testování hypotéz. Platí:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \approx \mathcal{N}(0, [\boldsymbol{J}(\boldsymbol{\theta})]^{-1}),$$

kde kovarianční maticí je inverzní matice k Fisherově informační matici.

Fisherova informační matice hraje významnou roli v teorii MLE, protože představuje míru informace o parametru $\boldsymbol{\theta}$, která je obsažena v každé proměnné. Prvky této matice jsou získány z kovariance nebo očekávané hodnoty druhých parciálních derivací logaritmické funkce pravděpodobnosti:

$$\begin{aligned} \boldsymbol{J}_{i,j}(\boldsymbol{\theta}) &= cov\left(\frac{\partial \ln(p(\boldsymbol{Y}|\boldsymbol{\theta}))}{\partial \theta_i}, \frac{\partial \ln(p(\boldsymbol{Y}|\boldsymbol{\theta}))}{\partial \theta_j}\right) \\ &= E\left(\frac{\partial \ln(p(\boldsymbol{Y}|\boldsymbol{\theta}))}{\partial \theta_i}, \frac{\partial \ln(p(\boldsymbol{Y}|\boldsymbol{\theta}))}{\partial \theta_j}\right) \end{aligned} \tag{2.4}$$

2.3 Metoda maximální věrohodnosti pro výpočet koeficientů v logistické regresi

Poté, co jsme se seznámili s principy metody maximální věrohodnosti a jejími příhodnými vlastnostmi, zaměříme se na její aplikaci v rámci logistické regrese.

Začneme konstrukcí sdružené pravděpodobnostní funkce. Předpokládáme, že výběr $\boldsymbol{Y}$ sleduje alternativní rozdělení s parametrem π, což značí prav-

děpodobnost úspěchu. Pravděpodobnostní funkce pro jednotlivou náhodnou veličinu Y_i je potom založena na následující formuli:

$$P(Y_i = y_i) = \pi^{y_i}(1-\pi)^{1-y_i} \quad y_i \in \{0,1\}, \quad i = 1, \ldots, n. \tag{2.5}$$

Díky předpokladu nezávislosti mezi náhodnými veličinami Y_i je možné vytvořit sdruženou pravděpodobnostní funkci jako součin individuálních pravděpodobnostních funkcí jednotlivých pozorování. To nám umožňuje modelovat pravděpodobnost celého souboru dat s ohledem na parametr π, který chceme pomocí metody maximální věrohodnosti odhadnout.

$$p(y_1, \ldots, y_n) = \prod_{i=1}^{n} \pi^{y_i}(1-\pi)^{1-y_i}. \tag{2.6}$$

Nyní budeme brát v úvahu podmíněnou pravděpodobnost ve vztahu k realizacím nezávislých proměnných x_i, jelikož se pravděpodobnost $\pi_i(\boldsymbol{\beta})$, že Y_i nabude hodnoty 1 za předpokladu hodnot x_i, bude podle naměřených dat měnit. Tato pravděpodobnost (1.10), reprezentovaná pomocí logistické funkce, se vypočítá takto:

$$\pi_i(\boldsymbol{\beta}) = P(Y_i = 1 | \boldsymbol{X_i} = \boldsymbol{x_i}) = \frac{e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}}.$$

Analogicky, pravděpodobnost $1 - \pi_i(\boldsymbol{\beta})$, že Y_i nabude hodnoty 0 za stejných podmínek, je dána vztahem:

$$1 - \pi_i(\boldsymbol{\beta}) = P(Y_i = 0 | \boldsymbol{X_i} = \boldsymbol{x_i}) = \frac{1}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}}.$$

Následně tyto pravděpodobnosti dosadíme do sdružené pravděpodobnostní funkce. Jelikož hodnoty y_i máme v datech, můžeme zaměnit jejich roli proměnných s koeficienty $\boldsymbol{\beta}$, čímž získáme věrohodnostní funkci ve tvaru:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}} \right)^{y_i} \left(\frac{1}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}} \right)^{1-y_i} = \prod_{i=1}^{n} \frac{(e^{\boldsymbol{\beta}^T \boldsymbol{x}_i})^{y_i}}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}}. \tag{2.7}$$

Navíc přejdeme na logaritmický tvar věrohodnostní funkce:

$$\ln L(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) = \sum_{i=1}^{n} y_i (\boldsymbol{\beta}^T \boldsymbol{x}_i) - \sum_{i=1}^{n} \ln(1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}). \tag{2.8}$$

Abychom přešli na systém věrohodnostních rovnic, je třeba, abychom funkci $l(\boldsymbol{\beta})$ zderivovali podle všech proměnných β_j:

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n} y_i x_{ij} - \frac{x_{ij} e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}} \quad j = 0, 1, \ldots, m. \tag{2.9}$$

Vytknutím x_{ij} z obou členů a položením do rovnosti s nulou dostáváme rovnice ve tvaru:

$$\sum_{i=1}^{n} \left[y_i - \frac{e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}} \right] x_{ij} = 0 \quad j = 0, 1, \ldots, m. \tag{2.10}$$

Tato soustava rovnic obsahuje $m + 1$ rovnic pro $m + 1$ neznámých. Řešení této soustavy se hledá iteračně.

Tvar (2.10) nám umožňuje formulovat problém v maticové formě, což usnadňuje výpočty, zejména při použití numerických metod pro iterativní hledání řešení. Maticový tvar vyjádříme jako:

$$\mathbf{0} = \boldsymbol{X}^T (\boldsymbol{Y} - \boldsymbol{\pi}(\boldsymbol{\beta})), \tag{2.11}$$

kde $\boldsymbol{X}_{n \times m}$ představuje matici nezávislých proměnných x_{ij} z našich dat, $\boldsymbol{Y}_{n \times 1}$ je vektor o závislých hodnot y_i, a $\boldsymbol{\pi}(\boldsymbol{\beta})_{n \times 1}$ je vektor podmíněných pravděpodobností $\pi_i(\boldsymbol{\beta}) = \frac{e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}}{1 + e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}}$ pro každé pozorování.

Nyní předpokládejme, že jsme řešení našli. Nezbývá než ověřit, zda nalezené hodnoty představují maximum. z matematické analýzy víme, že k takovému závěru dojdeme z negativní definitnosti Hessovy matice, která představuje matici druhých parciálních derivací. Využijeme Fisherovy informační matice, kterou jsme si již představili. (2.4)

Základem bude najít její odhad $\boldsymbol{J}(\boldsymbol{\beta})$. v praxi k tomu dospějeme tak, že vypočítáme matici druhých parciálních derivací $\left[\frac{\partial^2 l(\beta)}{\partial\beta_r\partial\beta_s}\right]_{r,s=0}^{m}$. Tuto matici vyčíslíme v bodě odhadu $\hat{\beta}_i$. Budeme tedy derivovat výraz na pravé straně z (2.10):

$$\begin{aligned}\frac{\partial^2 l(\beta)}{\partial\beta_r\partial\beta_s} &= \frac{\partial}{\partial\beta_s}\left(\sum_{i=1}^{n}\left[y_i - \frac{e^{\boldsymbol{\beta}^T\boldsymbol{x}_i}}{1+e^{\boldsymbol{\beta}^T\boldsymbol{x}_i}}\right]x_{ir}\right)\\ = -\sum_{i=1}^{n}\left(\frac{e^{\boldsymbol{\beta}^T\boldsymbol{x}_i}}{1+e^{\boldsymbol{\beta}^T\boldsymbol{x}_i}}\right)&\left(\frac{1}{1+e^{\boldsymbol{\beta}^T\boldsymbol{x}_i}}\right)x_{ir}x_{is} = \boldsymbol{H}(\boldsymbol{\beta}).\end{aligned} \tag{2.12}$$

Je důležité si uvědomit, že jsme právě vypočítali Hessovu matici. Pro přechod k Fisherově informační matici stačí vzít opačné hodnoty a dosadit odhady $\hat{\boldsymbol{\beta}}$:

$$\boldsymbol{J}(\hat{\boldsymbol{\beta}}) = -\boldsymbol{H}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{n}\left(\frac{e^{\boldsymbol{\beta}^T\boldsymbol{x}_i}}{1+e^{\hat{\boldsymbol{\beta}}^T\boldsymbol{x}_i}}\right)\left(\frac{1}{1+e^{\hat{\boldsymbol{\beta}}^T\boldsymbol{x}_i}}\right)x_{ir}x_{is}.$$

Uvedené výrazy v závorce jsme již dříve identifikovali jako pravděpodobnosti $\pi_i(\hat{\boldsymbol{\beta}})$ a $1-\pi_i(\hat{\boldsymbol{\beta}})$. To nám umožňuje zapsat Fisherovu informační matici v maticové formě:

$$\boldsymbol{J} = \boldsymbol{X}^T\boldsymbol{U}\boldsymbol{X},$$

kde $\boldsymbol{U} = \mathrm{diag}(\pi_1(\hat{\boldsymbol{\beta}})(1-\pi_1(\hat{\boldsymbol{\beta}})),\ldots,\pi_n(\hat{\boldsymbol{\beta}})(1-\pi_n(\hat{\boldsymbol{\beta}})))$. Podmínka pro pozitivní definitnost matice $\boldsymbol{A}$ je $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$ pro všechny $\boldsymbol{x}\in\mathbb{R}^m, \boldsymbol{x}\neq\boldsymbol{0}$, což

platí pro Fisherovu informační matici. To je způsobeno tím, že matice $\boldsymbol{U}$ se skládá z kladných pravděpodobností. Vzhledem k vztahu $\boldsymbol{J} = -\boldsymbol{H}$ dále plyne, že Hessova matice je negativně definitní, což potvrzuje, že nalezené řešení skutečně odpovídá maximu logaritmické věrohodnostní funkce.

3 Hledání vhodného modelu

Již víme, jak sestavit logistický model a vypočítat koeficienty β. Nyní si však představíme metody, které nám mohou pomoci rozhodnou, jaký model se k našim datům bude hodit nejvíce. Především se budeme snažit vyhnout tak zvanému overfittingu, který nastává, když zvolíme příliš mnoho nezávislých proměnných. Takový model může velmi dobře popisovat tréninková data, ale pro jiná data bude nepoužitelný. Snažíme se tedy o nalezení rovnováhy mezi vysvětlenou variabilitou a počtem použitých proměnných, abychom minimalizovali riziko overfittingu a zajistili, že model bude obsahovat pouze proměnné klíčové pro co nejpřesnější predikce.

3.1 Waldův test

Máme-li sestavený model a odhadnuté koeficienty beta, můžeme se podívat, které proměnné jsou v tomto modelu statisticky významné. To znamená, které proměnné mají relevantní příspěvky do modelu. Budou nás tedy zajímat hodnoty jednotlivých koeficientů β. Kdyby totiž některý byl roven 0, potom by příslušná nezávislá proměnná neměla žádný efekt na odhadovanou podmíněnou pravděpodobnost nastání jevu.

Waldův test přesně plní tuto úlohu a hypotézy jsou vzneseny tyto:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0.$$

Nulovou hypotézu zamítáme na hladině významnosti α, která činí zpravidla 0,05 (5 %) a určuje maximální pravděpodobnost falešného zamítnutí platné nulové hypotézy, tj. pravděpodobnost chyby prvního druhu.

Pro testování hypotézy použijeme statistiku:

$$Z_j = \frac{\hat{\beta}_i}{\sqrt{\widehat{var}(\hat{\beta}_i)}} = \frac{\hat{\beta}_i}{\widehat{SE}(\hat{\beta}_i)} \approx N(0,1). \tag{3.1}$$

Můžeme se však setkat s tvarem statistiky, který by byl v druhé mocnině, tudíž Z_j^2. v tomto případě by byla statistika přibližně z χ^2 rozdělení s jedním stupněm volnosti. Tato úprava nám může zajistit lepší výsledek, není-li naplněna normalita z předešlé statistiky. Zároveň je robustnější, tj. odolnější vůči odlehlým hodnotám

V naší situaci budeme porovnávat hodnotu koeficientu s jeho standardizovanou chybou, kterou vypočítáme takto:

$$\widehat{SE}(\hat{\beta}_j) = \sqrt{\widehat{var}(\hat{\beta}_j)} = \sqrt{(\mathbf{J}(\hat{\beta}_j)^{-1})_{jj}}.$$

Porovnáváme, kolikrát je hodnota koeficientu větší, než jeho chyba a kýženým výsledkem je hodnota statistiky $Z_j > 1$. Pokud by hodnota statistiky Z_j byla menší než 1, znamená to, že intervalový odhad koeficientu β_j zahrnuje 0. To naznačuje, že data nevyvrací možnost, že by j-tá proměnná neměla žádný vliv na hodnotu závislé proměnné. Pro rozhodnutí o platnosti hypotézy se podíváme na kritické hodnoty, kterými v tomto případě jsou kvantily normálního rozdělení $u_{1-\alpha/2}$ a $-u_{1-\alpha/2}$. Vyskytuje-li se statistika v intervalu mezi těmito hodnotami, není možné nulovou hypotézu zamítnout. v opačném případě jsme však dostatečně přesvědčení o tom, že se koeficient nule nerovná. Pomocí hodnot, které se v tomto testu vyskytovaly, můžeme vypočítat intervalový odhad koeficientů na hladině spolehlivosti $1-\alpha$:

$$\left\langle \hat{\beta}_j - u_{1-\alpha/2}\sqrt{\widehat{var}(\hat{\beta}_j)}; \hat{\beta}_j + u_{1-\alpha/2}\sqrt{\widehat{var}(\hat{\beta}_j)} \right\rangle. \tag{3.2}$$

Uveďme si nyní příklad, aby bylo zřejmé, jak testování funguje. Předpokládejme, že jsme vypočítali hodnoty $\hat{\beta}_j = 0{,}5$ a $\sqrt{\widehat{var}(\hat{\beta}_j)} = 0{,}7$. z těchto hodnot a vztahu (3.1) můžeme vypočítat statistiku Z_j, která vyjde přibližně 0,714. Při hladině významnosti $\alpha = 0{,}05$ je kritický obor, který získáme z˘ kvantilů z normálního rozdělení, $\langle -1{,}645; 1{,}645 \rangle$. Zde můžeme vidět, že by koeficient měl být více než 1,645krát větší než jeho směrodatná odchylka, aby byla nulová hypotéza zamítnuta. v našem případě tento požadavek není splněn a hypotéza o nulovosti koeficientu β_j nelze zamítnout. Podívejme se, jak by vypadal intervalový odhad tohoto koeficientu. Po dosazení do vztahu (3.2) nám vyjde interval $\langle -0{,}6515; 1{,}6515 \rangle$. Je zřejmé, že hodnota koeficientu β_j se může rovnat nule, jelikož intervalový odhad tento bod obsahuje.

Na závěr je nutné podotknout, že tato metoda má svá omezení. Jedná se o parametrický test, tudíž předpokládáme, že s rostoucím rozsahem dat, se bude statistika přibližovat danému rozdělení. v tomto případě to bylo $N(0, 1)$. Pokud však nemáme dostatek dat, nemůžeme zaručit, že statistika bude skutečně následovat toto rozdělení. Finální rozhodnutí tedy nemusí být správné.

3.2 Informační kritérium AIC

Informační kritéria, jako je například AIC (Akaikeho informační kritérium), slouží k porovnávání různých statistických modelů. Tato kritéria zohledňují, jak dobře příslušné modely reprezentují data a zároveň modely penalizují za přílišnou složitost, čímž se předchází overfittingu.

Obecně platí, že čím více parametrů pro model zvolíme, tím lépe bude na data pasovat. Cílem statistiky však není najít model, který perfektně pasuje na trénovací data, ale snažíme se informaci zobecnit. Data jsou tedy jen vzorkem z mnohem většího celku a když bychom vzali jiný vzorek, tak

očekáváme že se bude od původního lišit. v případě overfittingu by příliš přesný model nebyl vůbec použitelný pro jiné vzorky.

S AIC se můžeme setkat hned v summary modelu `glm` funkce ve statistickém softwaru R. Jeho výpočet je založen na počtu odhadnutých parametrů modelu (k) a hodnoty věrohodnostní funkce ($\hat{L}$) po její maximalizaci:

$$AIC = 2k - ln(\hat{L}).$$

Za předpokladu, že máme p modelů a ke každému máme vypočítané hodnoty kritéria $AIC_1, AIC_2, \ldots, AIC_p$, potom model s nejnižší hodnotou kritéria představuje nejlepší kompromis mezi jeho složitostí a přesností predikce.

3.3 Pseudokoeficienty determinace

Pseudokoeficienty determinace slouží jako alternativy k tradičnímu koeficientu determinace R^2 z lineární regrese, avšak jsou určené pro nelineární modely, jako je logistická regrese. Ačkoliv mají podobný cíl jako R^2, a to popsat míru variability v závislé proměnné, kterou model vysvětluje, jejich interpretace je často méně intuitivní a mohou být nejednoznačné. Proto knihovna `pcls` v R nabízí hned tři varianty pseudokoeficientů pro hlubší pochopení. Všechny tyto pseudokoeficienty vychází z věrohodnostních funkcí prázdného modelu (L_{null}), kde je jen intercept, a odhadnutého modelu (L_M). [9, 19]

- McFaddenovo R^2_{Mc} :

$$R^2_{Mc} = 1 - \frac{ln(L_M)}{ln(L_{null})}$$

Tento koeficient se snaží kvantifikovat, jak moc se model zlepší ve srovnání s prázdným modelem. Výhodou je jednoduchost výpočtu a snadná interpre-

tace, ovšem koeficient má tendenci nadhodnocovat vysvětlenou variabilitu modelů konstruovaných na vyšším počtu pozorování.

- Cox-Snellovo R^2_{CS} :

$$R^2_{CS} = 1 - \left(\frac{L_M}{L_{null}}\right)^{2/n}$$

Jedná se o modifikovanou verzi McFaddenova koeficientu, která řeší nadhodnocování. Nevýhodou však je, že ani perfektní model, který má věrohodnostní funkci rovnou 1, nedosáhne maximální hodnoty 1, nýbrž $1 - (L_{null})^{2/n}$. Tato vlastnost značně komplikuje jeho interpretaci.

- Nagalkerkeho R^2_{Ng} :

$$R^2_{Ng} = \frac{1 - (\frac{L_{null}}{L_M})^{2/n}}{1 - (L_{null})^{2/n}}$$

Nagelkerkeho koeficient je Cox-Snellův koeficient vydělený jeho maximální hodnotou, díky čemuž může dosáhnout hodnoty 1, což usnadňuje jeho interpretaci a připodobňuje ho jeho protějšku z lineární regrese. Nevýhodou je, že může být vnímán jako umělé vylepšení skutečné hodnoty.

Všechny tyto koeficienty mají nevýhodu, že upřednostňují modely s vyšším počtem parametrů, což může vést k riziku overfittingu. Při vyhodnocování modelů je proto důležité brát v úvahu i další kritéria a metriky.

4 Hodnocení modelů binární klasifikace

4.1 Klasifikační tabulka

Klasifikační tabulka [4] vychází z předpokladu, že před modelováním byla data rozdělena na trénovací a testovací sadu. Tedy n pozorování jsme rozdělili na k testovacích a $n-k$ trénovacích. Modelování probíhá na trénovací sadě, která bývá zpravidla větší a využívá informace, do které skupiny pozorování patří. Jestliže jsme našli model M, pro všech k testovacích pozorováních si vypočítáme logistické skóre.

$$s_i = \hat{\pi}_i(\hat{\boldsymbol{\beta}}) = \frac{e^{\hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i}}{1 + e^{\hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i}}, \quad i = 1, \ldots, k. \tag{4.1}$$

Na základě tohoto skóru a vhodného prahového bodu (cut-off) C_M provedeme klasifikaci, tudíž rozdělíme pozorování do kategorií závislé proměnné $\{0, 1\}$. Jestliže bude $s_i \geq C_m$, budeme předpokládat, že patří do kategorie 1. v opačném případě kdy $s_i < C_m$ bude pozorování náležet do kategorie 0. V testovacích datech je nám skutečná příslušnost do skupin známá, proto si můžeme v klasifikační tabulce (viz tabulka 1) zobrazit, jak úspěšně jsme klasifikovali.

	predikovaná 0	**predikovaná 1**
skutečná 0	#Skutečně negativních (**SN**)	#Falešně pozitivních (**FP**)
skutečná 1	#Falešně negativních (**FN**)	#Skutečně pozitivních (**SP**)

Tabulka 1: Klasifikační tabulka

Díky klasifikační tabulce můžeme vypočítat různé metriky. My si zde uvedeme tři základní metriky. První je celková přesnost predikce (accuracy), která poskytuje přehled o celkové úspěšnosti modelu při klasifikaci. Jedná se o počet správně klasifikovaných pozorování, vydělený celkovým počtem

pozorování k, tudíž pravděpodobnost správného zařazení:

$$accuracy = \frac{\mathbf{SN} + \mathbf{SP}}{\mathbf{SN} + \mathbf{SP} + \mathbf{FN} + \mathbf{FP}}$$

Další metrikou je senzitivita, která představuje pravděpodobnost, že model správně identifikuje pozitivní pozorování. Jinými slovy udává, jak dobře model rozpozná skutečně pozitivní případy. Počítá se jako poměr správně klasifikovaných pozitivních pozorování k celkovému počtu skutečně pozitivních pozorování:

$$sensitivity = \frac{\mathbf{SP}}{\mathbf{SP} + \mathbf{FN}}$$

Protějškem k senzitivitě je specificita, která se počítá jako poměr správně klasifikovaných negativních pozorování k celkovému počtu skutečných negativních pozorování. Specificita představuje pravděpodobnost správného zařazení, jestliže bylo pozorování ve skutečnosti negativní:

$$specificity = \frac{\mathbf{SN}}{\mathbf{SN} + \mathbf{FP}}$$

Celková přesnost, senzitivita a specificita poskytují komplexní přehled o výkonu klasifikačního modelu a umožňují nám posoudit jeho efektivitu při různých scénářích. Celková přesnost nám dává obecný přehled o úspěšnosti modelu, ale může být zavádějící v případech s nevyváženými daty, kde jedna kategorie výrazně převažuje. Senzitivita a specificita jsou proto klíčové pro pochopení, jak model zvládá správně identifikovat pozitivní a negativní případy. v praxi se často používají všechny tři metriky společně, aby poskytly úplný obrázek o výkonnosti modelu a umožnily lepší rozhodování při úpravách modelu, či volbě prahové hodnoty.

4.2 ROC křivka

ROC (Receiver Operating Characteristic) křivka [4] se používá k hodnocení a vizualizaci výkonnosti klasifikačních modelů. Ukazuje, jak se mění specificita a senzitivita v závislosti na volbě prahového bodu. Před konstrukcí si zaveďme indikátorovou funkci

$$I(A) = \begin{cases} 1 & \text{pokud a platí} \\ 0 & \text{jinak} \end{cases} \tag{4.2}$$

a na jejím základě definujeme funkce:

$$F_0(x) = \frac{1}{n_0} \sum_{i=1}^{n} I(s_i < x, y_i = 0), \quad x \in (0,1),$$

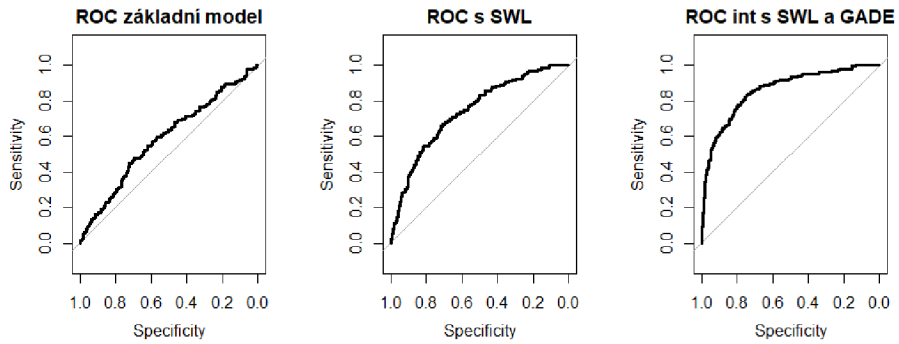
$$F_1(x) = \frac{1}{n_1} \sum_{i=1}^{n} I(s_i \geq x, y_i = 1), \quad x \in (0,1),$$

kde n_0 představuje počet pozorování, pro která platí, že $y_i = 0$ a n_1 počet pozorování, kde $y_i = 1$. Funkce F_0 tedy představuje pravděpodobnost, že skóre z modelu s_i je menší jako prahový bod x (čímž je pozorování klasifikováno jako 0) a zároveň, že skutečná hodnota se 0 rovnala. Mluvíme tedy o specificitě. F_1 představuje pravděpodobnost, že skóre s_i je větší rovno jako prahový bod x (pozorování je klasifikováno jako 1) a že skutečná hodnota se rovnala 1, a tak je zase definována senzitivita. Na základě těchto funkcí můžeme definovat ROC křivku jako množinu dvojic senzitivity a specificity:

$$\{(F_0(x), F_1(x)); x = C_m \in (0,1)\}.$$

Tyto body si můžeme zobrazit. Pro představu o tom, jak má ROC křivka vypadat, si uvedeme příklady ROC křivek u některých modelů úzkostných

poruch, se kterými se setkáme v praktické části.



Obrázek 1: ROC křivky nejjednoduššího modelu (vlevo), modelu s pomocnou proměnnou (uprostřed) a nejkomplexnějšího modelu (vpravo)

V ideálním případě bychom chtěli, aby se senzitivita i specificita rovnaly jedné. Tomuto scénáři se však můžeme jen přiblížit. Čím více se křivka opticky přibližuje bodu $(1, 1)$, tím lepší má model klasifikační schopnosti. Když se však model přibližuje úhlopříčce z bodu (1,0) do bodu (0,1), tak má model podobné vlastnosti, jako kdybychom klasifikovali náhodně. Na obrázku 1 můžeme vidět, jak se s rostoucí komplexitou modelu ROC křivka zaobluje směrem k bodu (1,1). Čím blíže by tomuto bodu byla, tím lepší klasifikační schopnosti od modelu očekáváme.

Na základě ROC křivky můžeme také určit optimální hraniční bod C_m, například pomocí Youdenova indexu, jehož maximalizací získáme hraniční bod, který současně zajišťuje co nejvyšší senzitivitu (sen) i specificitu ($spec$):

$$J_{MAX} = \max_{C_m}\{sen(C_m) + spec(C_m) + 1\}$$

Díky ROC můžeme i číselně vyjádřit plochu pod křivkou. Takovéto kritérium se nazývá AUC (Area under curve) a vyjadřuje, jak dobrou má model schopnost rozlišovat třídy a zároveň jeho robustnost vůči volbě prahové

hodnoty. Jeho hodnota je interpretována jako pravděpodobnost, že náhodné pozitivní pozorování bude mít vyšší skóre přidělené modelem než náhodné negativní pozorování.

5 Praktická část

Data (viz Příloha č. 1) byla sesbírána v roce 2017 pracovníky psychologické katedry na mnichovské univerzitě Marianem Sauterem a Dejanem Draschkowem.[23] Účelem bylo hlubší prozkoumání vztahů mezi hraním her a psychickými problémy. Sběr probíhal prostřednictvím dotazníku, který byl vyvěšen na Redditu, konkrétně na komunitním serveru hry League of Legends.

Dotazník se skládal ze tří částí. První část zahrnovala psychologické inventáře, konkrétně na generalizovanou úzkostnou poruchu (GAD-7), sociální fobii (SPIN), spokojenost se životem (SWL) a narcismus (SINS). Druhá část se zabývala hracími návyky, jako je průměrný počet nahraných hodin za týden, preference her a motivace k hraní. Třetí část se zaměřovala na demografické informace, zahrnující věk, pohlaví, zemi původu a pracovní status osoby. Výzkumníkům se podařilo získat odpovědi od 13464 respondentů.

5.1 Psychometrické dotazníky

Ještě před tím než si dotazníky představíme, vysvětleme si princip vnitřní reliability (spolehlivost) testu[27], která má za cíl měřit, zda všechny položky v testu měří stejný koncept. Nejčastěji se vyjadřuje pomocí Cronbachovy alfy.:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} \sigma_{y_i}^2}{\sigma_y^2}\right),$$

kde k představuje počet položek v dotazníku, $\sigma_{y_i}^2$ rozptyl i-té položky dotazníku a σ_y^2 celkový rozptyl skórů. Hodnoty se pohybují v rozmezí od 0 do 1, přičemž hodnota nad 0,7 je považována za přijatelnou a hodnota nad 0,8 za dobrý indikátor vnitřní reliability. Pokud všechny položky měří stejný kon-

cept, potom by jejich skóre měla být korelovaná a suma rozptylů jednotlivých položek $\sigma^2_{y_i}$ výrazně nižší než celkový rozptyl $\sigma^2_{y_i}$.

5.2 GAD7 - generalizovaná úzkostná porucha

Generalizovaná úzkostná porucha (GAD - generalized anxiety disorder) se projevuje nekontrolovatelnými úzkostmi a obavami, které postiženého doprovázejí v každodenních činnostech. Tyto obavy jsou často nepřiměřené k daným situacím a mohou výrazně ovlivňovat život pacienta. Na rozdíl od běžných úzkostí, které zažívá většina lidí, se jedinec cítí úzkostně téměř trvale. Diagnóza této poruchy byla oficiálně zařazena do Diagnostického a statistického manuálu duševních poruch (DSM-III) v roce 1980. u pacientů je typická únava, častá návštěvnost doktorů skrze zdravotní problémy, podrážděnost a tenze. Tato porucha je častější u žen než u mužů. Diagnóza GAD je často spojena s depresí a vyšší tendencí k užívání návykových látek. [5]

Dotazník GAD-7 (viz Příloha č. 2) byl navrhnut až v roce 2006 a skládá se ze sedmi otázek. Respondenti jsou tázáni, jak často v posledních dvou týdnech prožívaly určité stavy. Odpovědi jsou na škále „vůbec“, „několik dnů“, „kolem poloviny dnů“ a “téměř každý den“. Výsledné skóre se získává sečtením odpovědí a může maximálně dosahovat 21 bodů. Skóry 5, 10 a 15 reprezentují hranice, kdy rozlišujeme mírnou, střední a závažnou intenzitu poruchy. Doporučená hranice na diagnostiku poruchy je však 10 a více. Při této hranici je senzitivita pro diagnostiku generalizované úzkostné poruchy 89 % , specificita 82 % a Cronbachovo alfa se pohybuje mezi 0,85 až 0,92, což naznačuje vysokou míru spolehlivosti. Dotazník je navíc poměrně vhodný i na diagnostiku panické poruchy, sociálně-úzkostné poruchy nebo posttraumatické stresové poruchy. [16] Proto budeme při interpretaci modelů uvažovat širší spektrum úzkostných poruch.

5.3 SPIN - sociálně-úzkostná porucha

Sociální fobie, respektive sociálně-úzkostná porucha, je jeden z nejběžnějších typů úzkostných poruch. Způsobuje emocionální i tělesné problémy při osobním kontaktu s lidmi, při telefonování nebo jezení na veřejnosti. Prohlubuje ji stydlivost a nízké sebevědomí. Porucha sama naopak prohlubuje predispozice k depresím nebo užívání návykových látek. Byla zařazena do manuálu duševních poruch roku 1980. Její definování se však vyvíjelo a momentálně čerpáme z definice z roku 2013 (DSM-V). [25]

Dotazník SPIN (viz Příloha č. 3) byl navržen v roce 2000. Skládá se ze sedmnácti prožitků, na které respondent reaguje podle toho, jak často se s nimi v posledním týdnu setkal. Odpovídá na škále „vůbec“, „trochu“, „poněkud“, “hodně“ a „extrémně“. Výsledné skóre vzniká součtem odpovědí a může maximálně dosahovat 68 bodů. Skóry 20, 31, 41 a 51 představují hranice pro slabou, mírnou, závažnou a velmi závažnou intenzitu sociální fóbie. Doporučeným mezníkem pro diagnózu je však 24, při kterém test dosahuje senzitivity 81,2 %, specificity 85,1 % a Cronbachova alfa se pohybuje v rozmezí 0,80 – 0,94, což naznačuje vysokou míru spolehlivosti. [21]

5.4 SINS - narcismus

Narcismus je osobnostní charakteristika, která se projevuje zvýšeným vnímáním vlastní důležitosti, potřebou obdivu a často nedostatkem empatie vůči ostatním. v určité míře může být narcismus považován za normální, a dokonce zdravý aspekt osobnosti. Motivuje jednotlivce k dosažení úspěchu, seberealizaci a buduje sebevědomí. Lidé s určitými narcistickými rysy nemusí nutně trpět poruchou osobnosti a mohou být schopni navazovat a udržovat zdravé vztahy.

Krajním případem je narcistická porucha osobnosti, která byla definovaná již v DSM-III v roce 1980. Je charakterizována extrémními narcistickými rysy. Osoby s touto poruchou vykazují vzorce grandiózního myšlení, potřebu neustálého obdivu, manipulativní chování a výrazný nedostatek empatie. Obvykle se však jedná o pouhou masku, která zakrývá vnitřní nejistotu a věčnou nespokojenost se sebou samým. Tyto charakteristiky jsou musí být pro diagnózu dostatečně silné, trvalé a významně narušující osobní, profesionální a sociální fungování jedince. [22]

Dotazník SINS (viz Příloha č. 4) byl navrhnut roku 2014, jako kratší alternativa k jiným dotazníkům zkoumajícím narcismus. Skládá se pouze z jedné položky, kterou je prohlášení „jsem narcis“. Respondent odpovídá na škále „rozhodně nesouhlasím“, „spíše nesouhlasím“, „napůl souhlasím“, „spíše souhlasím“, „rozhodně souhlasím“. Navzdory své jednoduchosti podává dotazník dobré výsledky, obzvlášť v online dotaznících. Jeho výsledky jsou významně korelovány s komplexnějšími testy (např. NPI) a zároveň nekorelují se sebereflexí respondentů. [12]

5.5 Metrika SWLS - spokojenost se životem

Spokojenost se životem je často měřeným faktorem, který upřesňuje, v jaké životní situaci se pacient momentálně nachází. Může hrát klíčovou roli v diagnostice mnoha poruch a pomáhat v predikci rizika pokusů o sebevraždu. Vyšší úroveň spokojenosti se životem je často spojována s lepším duševním zdravím a nižším rizikem deprese a úzkosti. Naopak nízká úroveň spokojenosti může signalizovat zvýšené riziko duševních problémů a potřebu intervence.

Dotazník SWLS (viz příloha č. 5) byl navrhnut v roce 1985. Skládá se z pěti prohlášení, na která respondenti reagují na škále 1-7 podle míry sou-

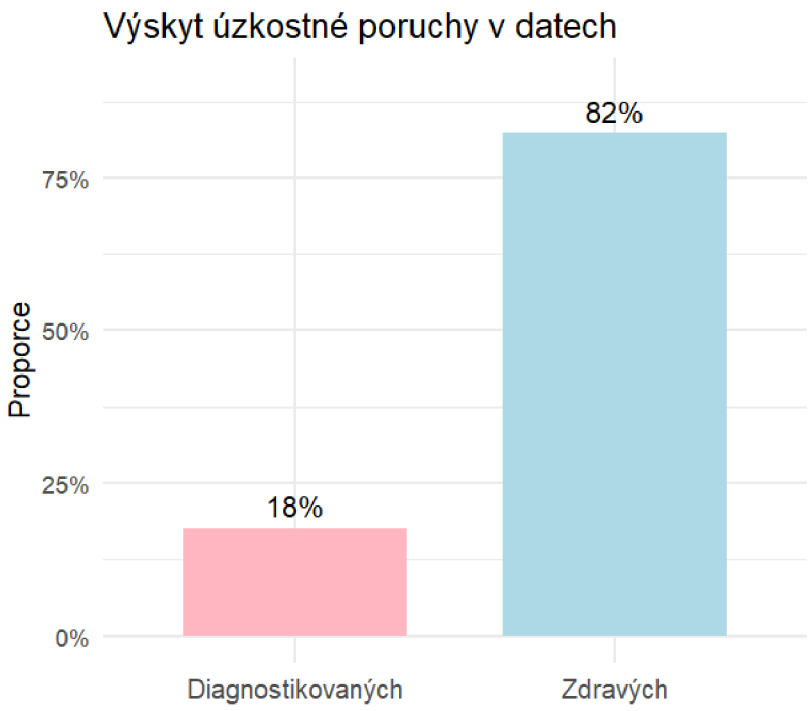
hlasu, tedy „silně nesouhlasím“, „nesouhlasím“, „spíše nesouhlasím“, „nevím“, „spíše souhlasím“, „souhlasím“, „nesouhlasím“. Výsledné skóre vzniká sečtením odpovědí, což znamená, že celkové skóre může dosahovat minimálně 5 bodů a maximálně 35 bodů. Obvykle se nepoužívají pevně definované hranice, místo toho se skóre považuje za kontinuální škálu. Pro potřeby analýzy však může být skóre kategorizováno jako: 5-9 extrémně nízká, 10-14 nízká, 15-19 podprůměrná, 20-24 průměrná, 25-29 vysoká a 30-35 jako extrémně vysoká spokojenost. Cronbachovo alfa se pohybuje v rozmezí 0,7-0,9, což naznačuje vysokou vnitřní reliabilitu a spolehlivost dotazníku při měření spokojenosti se životem. [6, 26]

5.6 Čištění dat a popisná statistika

V následujících kapitolách si ukážeme, jakým způsobem byla data čištěna a upravována pro tvorbu co nejlepších modelů. Následně se podíváme na popisnou statistiku očištěných dat. u některých pozorování chyběly informace, byly příliš odlehlé nebo hodnoty zastupovalo málo respondentů. Tato pozorování byla z původního datového souboru odebrána. Po čištění zbyly odpovědi od 10335 respondentů.

5.6.1 Rozdělení diagnóz na základě dotazníku GAD-7

Skóry dotazníku GAD-7 byly použity jako závislá proměnná v jednom z modelů, což vyžadovalo jejich transformaci na binární proměnnou. Jinými slovy bylo třeba „diagnostikovat“ přítomnost úzkostí. Jak už bylo zmíněno v části věnované dotazníku GAD-7, optimální hranicí diagnózy je hodnota skóru 10. u těch, kteří dosáhli alespoň této hodnoty nebo vyšší, budeme předpokládat přítomnost úzkostné poruchy. v opačném případě předpokládáme její nepřítomnost.

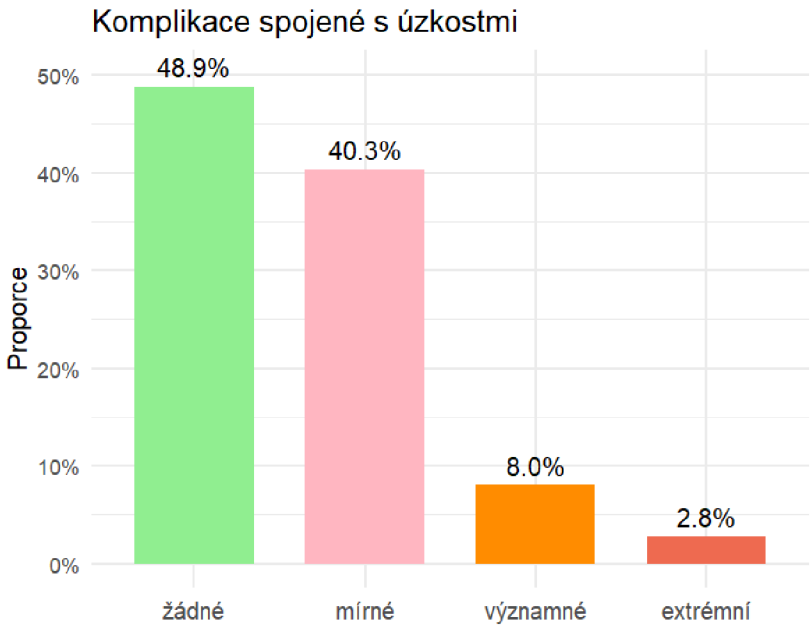


Obrázek 2: Proporční četnosti respondentů podle výskytu úzkostné poruchy diagnostikované na základě skórů GAD-7.

V grafu na obrázku 2 můžeme vidět, že by 18 % respondentů bylo diagnostikováno s úzkostnou poruchou. v roce 2021 byla odhadnuta prevalence úzkostných poruch 23 %, a to na vzorku dospívajících a mladých dospělých.[26] Jelikož se později dozvíme, že respondenti v datech se pohybují převážně ve věkovém rozmezí 18-21 let, můžeme se domnívat, že volba rozdělovacího kritéria byla vhodná.

Zároveň si při modelování budeme pomáhat doplňující odpovědí, která se do samotného skóre nezapočítává. Po zodpovězení otázek, týkajících se různých situací spojených s úzkostmi, byli respondenti tázáni, jak jim tyto situace komplikovaly každodenní činnosti. v grafu na obrázku 3 vidíme, že většina respondentů nemá žádné nebo jen mírné komplikace. Téměř 11 %

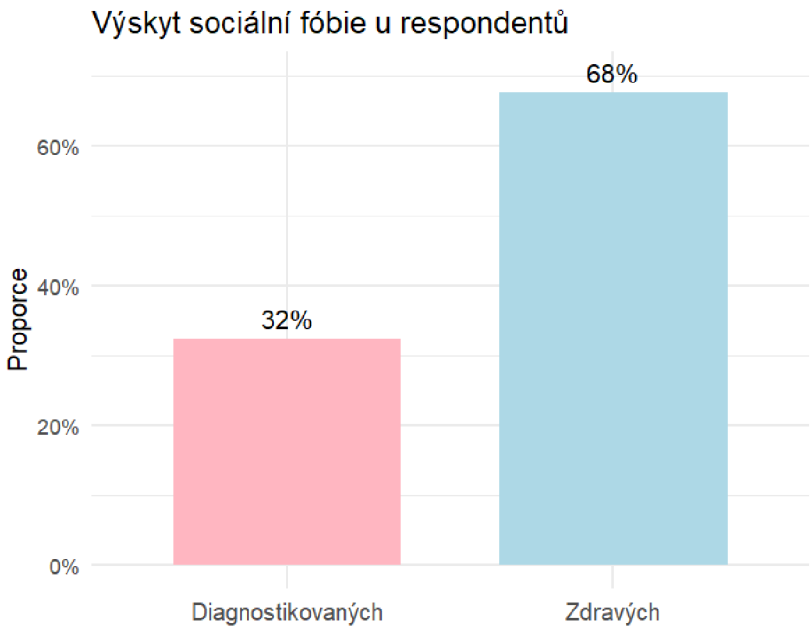
respondentů zažívá silné komplikace vlivem úzkostí.



Obrázek 3: Proporční četnosti respondentů na základě odpovědí týkajících se komplikací způsobených úzkostmi.

5.6.2 Rozdělení diagnóz na základě dotazníku SPIN

V druhém modelu je predikována pravděpodobnost sociální fóbie. Obdobně jako s GAD-7 bylo třeba respondenty rozdělit do dvou skupin podle dosaženého skóru. Opět jsme zvolili hraniční hodnotu, která by měla mít nejvyšší spolehlivost, a tou je skóre 24.

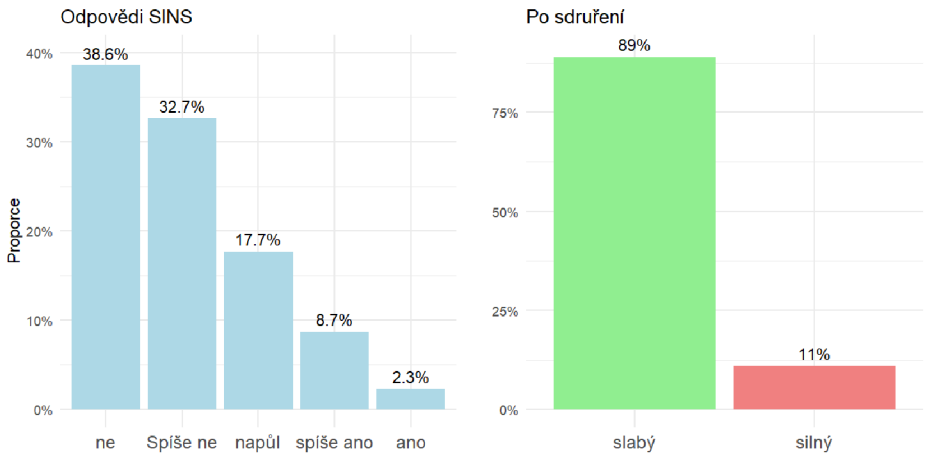


Obrázek 4: Proporční četnosti respondentů podle výskytu sociální fóbie, diagnostikované na základě dotazníku SPIN

Proporční hodnota diagnostikovaných v grafu na obrázku 4 činí 32 %. Pro porovnání byla v roce 2020 odhadovaná prevalence sociálně úzkostné poruchy v mladé populaci 36 %. [11]

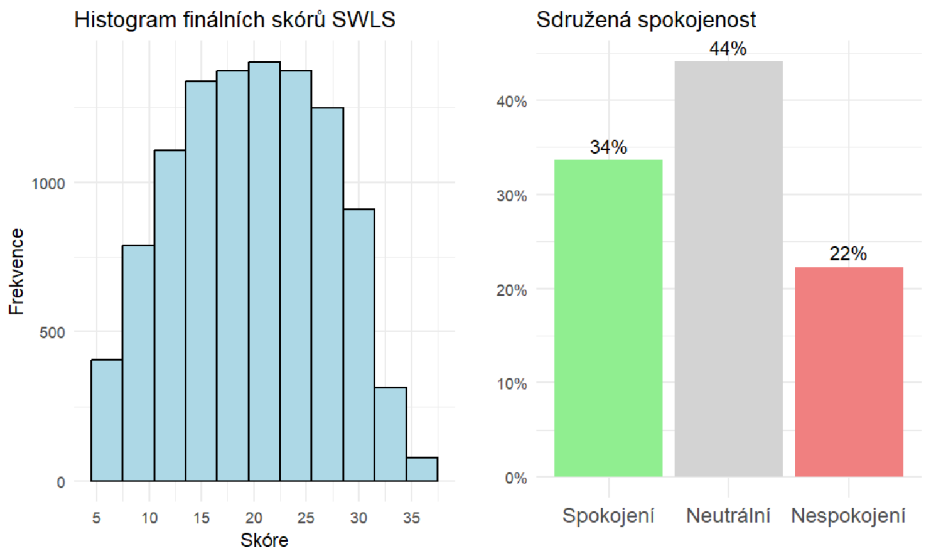
5.6.3 Ostatní psychometrické proměnné

Narcismus byl měřen pomocí dotazníku SINS, kde respondenti odpovídali na otázku, jak moc se s narcismem ztotožňují. Při modelování nebyl statisticky významný rozdíl mezi prvními třemi kategoriemi a u posledních dvou měl rozdíl stejný charakter. Proto byly vytvořeny pouze dvě kategorie, a to podle odhadované intenzity narcismu. v grafu na obrázku 5 můžeme vidět, že 11 % respondentů se s narcismem spíše ztotožňuje.



Obrázek 5: Proporční četnosti respondentů podle výsledků SINS (vlevo) a po sdružení do nových kategorií (vpravo).

U spokojenosti se životem byly vypočítány skóry z dotazníku SWLS. Kvůli lepší interpretaci byly nejdříve respondenti zařazeni do šesti kategorií, které jsme definovali v kapitole o SWLS. Při modelování nebyl signifikantní rozdíl v sousedních kategoriích, proto byly redukovány na. Po přeskupení reprezentuje kategorie „Neutrální“ respondenty, kteří měli skór mezi 16-24 body. Respondenti, kteří měli skór vyšší než 24 jsou označeni jako „Spokojení“ a ti co měli skóre nižší než 16 jsou označeni jako „Nespokojeni“.



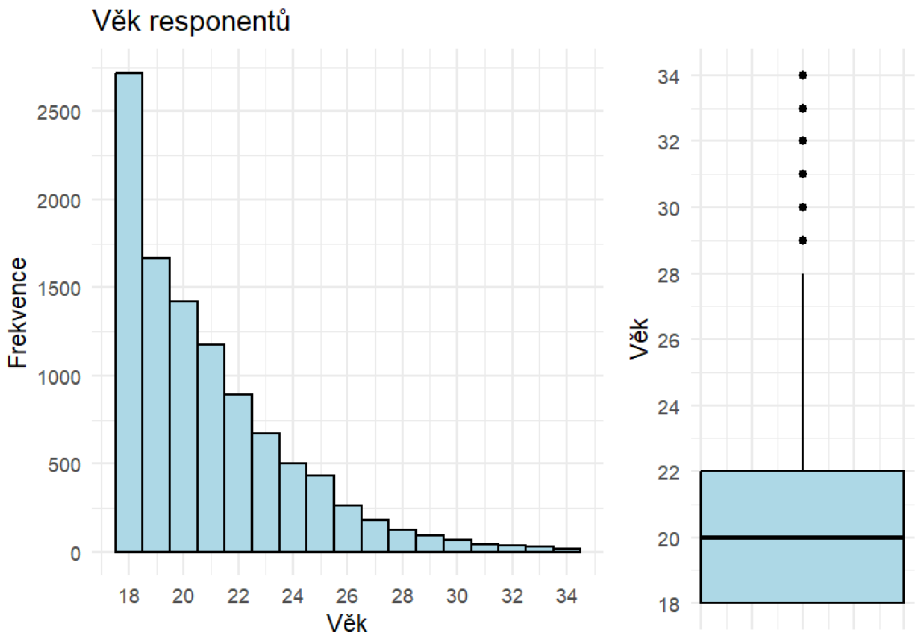
Obrázek 6: Histogram respondentů podle dosaženého skóru SWLS (vlevo) a proporční četnosti po sdružení do kategorií (vpravo).

V grafu na obrázku 6 se nejvyšší podíl respondentů pohybuje na neutrální škále spokojenosti. Výrazně spokojených respondentů je více než výrazně nespokojených.

5.6.4 Demografické informace

Mezi demografickými informacemi se ukázaly jako statisticky významné tyto proměnné: věk, pohlaví, nejvyšší dosažené vzdělání a země, ze které respondent pochází. Další proměnnou v datech byl pracovní status. Ve finálním modelu však nefiguruje.

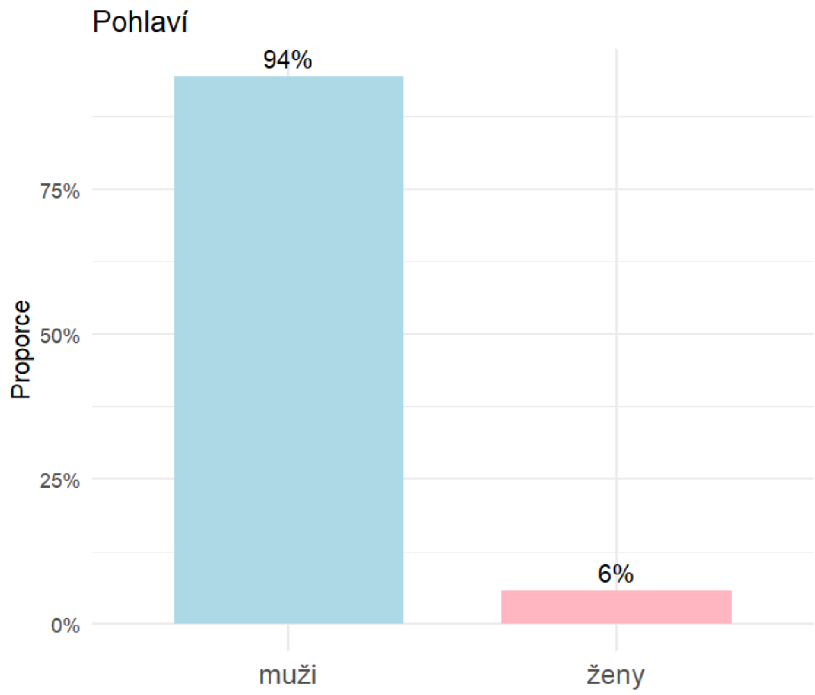
Převážná většina respondentů byla ve věku okolo dvaceti let. Nejstarší respondent měl 63 let, avšak takových výjimek bylo velmi málo, a proto je považujeme za odlehlá pozorování, jelikož by mohly zkreslovat výsledky. Pro detekci odlehlých hodnot byl použit trojnásobek mezikvartilového rozpětí (IQR), čímž byla stanovena věková hranice 34 let.



Obrázek 7: Histogram respondentů dle jejich věku (vlevo) a krabicový graf (vpravo).

Z krabicového grafu na obrázku 7 vyplývá, že 75 % respondentů je ve věku mezi 18 a 22 lety. Medián činí 20 let a horní fousek končí u 28 let. Vyšší hodnoty jsou v datech zastoupeny pouze v desítkách případů, což lze vidět i na histogramu, jelikož s rostoucím věkem pozorování ubývá.

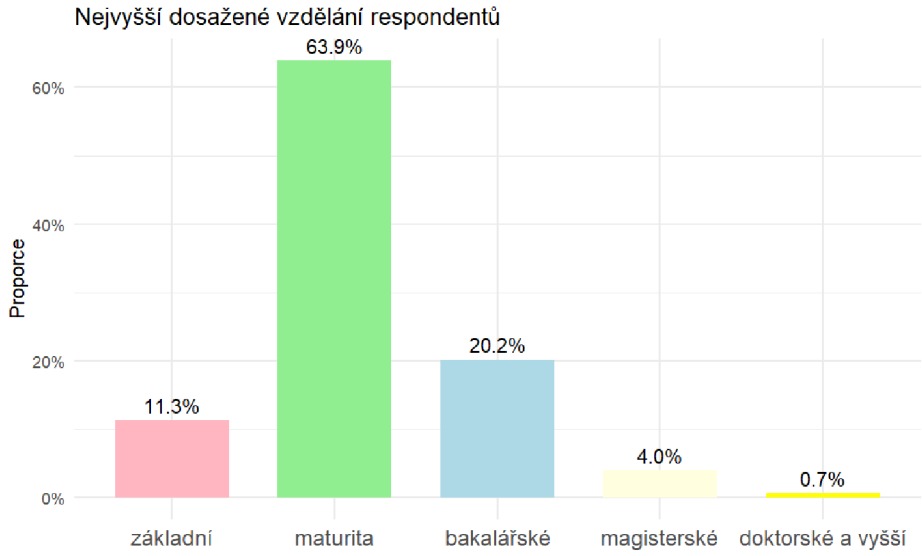
U pohlaví měli respondenti možnost kromě odpovědi muž/žena zadat i kategorii „jiné“. Tuto možnost vybralo 52 respondentů. Kvůli nízkému počtu těchto pozorování a složité interpretaci byla tato pozorování odebrána.



Obrázek 8: Proporční četnosti respondentů podle jejich pohlaví.

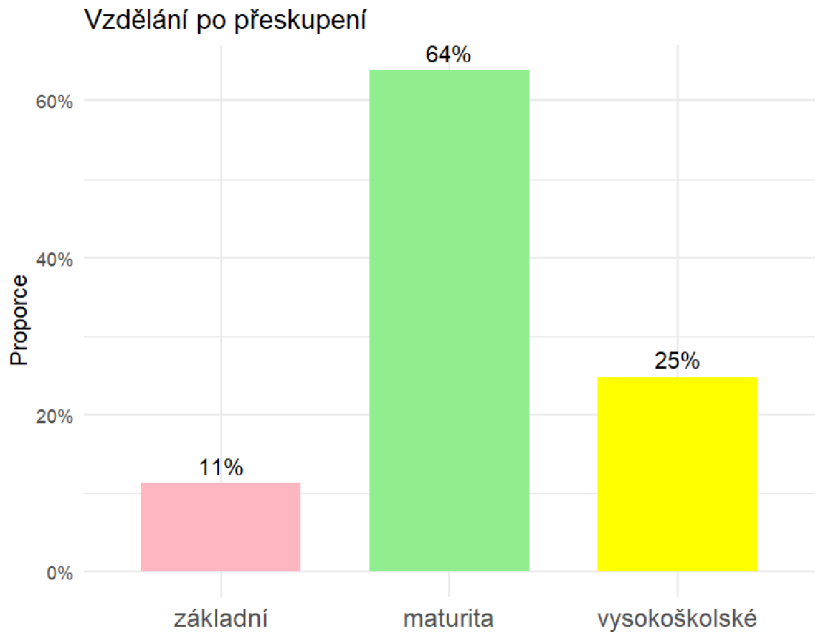
V grafu na obrázku 8 vidíme, že převážná většina respondentů byli muži. Žen se v datech vyskytovalo 6 %, přesněji 586 respondentek.

Další proměnnou bylo nejvyšší dosažené vzdělání. Budeme se na něj odkazovat pomocí českých ekvivalentů, ačkoliv byl dotazník mezinárodní.



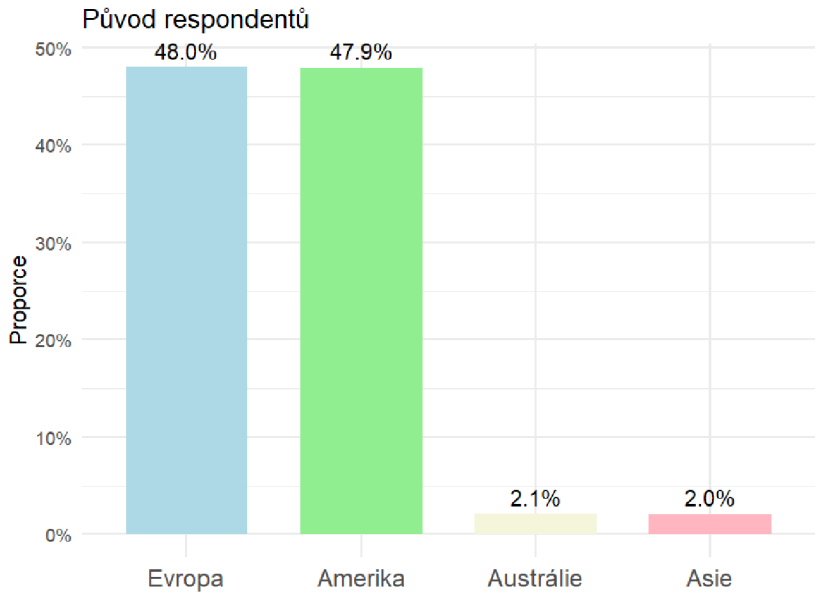
Obrázek 9: Proporční četnosti respondentů podle jejich nejvyššího dosaženého vzdělání.

Vzhledem k tomu, že spousta respondentů měla kolem 18 let, není překvapivé, že většina respondentů dosáhla nanejvýš maturitního vzdělání, viz obrázek 9. Protože se v modelech neprokázal statisticky významný rozdíl mezi vysokoškolskými tituly, byly sdruženy do jedné proměnné.



Obrázek 10: Proporční četnosti vzdělání respondentů po přeskupení.

Respondenti byli také tázáni, ze které země pochází. v datech byly uvedeny zkratky ve formátu ISO3. Na základě těchto zkratek byly státy sdruženy do proměnných reprezentující kontinenty. z Afriky pocházelo pouze 40 respondentů, proto byla z modelování vyřazena.



Obrázek 11: Proporční četnosti respondentů podle kontinentu, ze kterého pochází.

Na obrázku 11 vidíme, že převážná většina respondentů pochází z evropských a amerických zemí, pravděpodobně proto, že byl dotazník psán anglicky a zveřejněn na anglicky mluvícím serveru. Austrálie je zastoupena 219 respondenty a Asie 210.

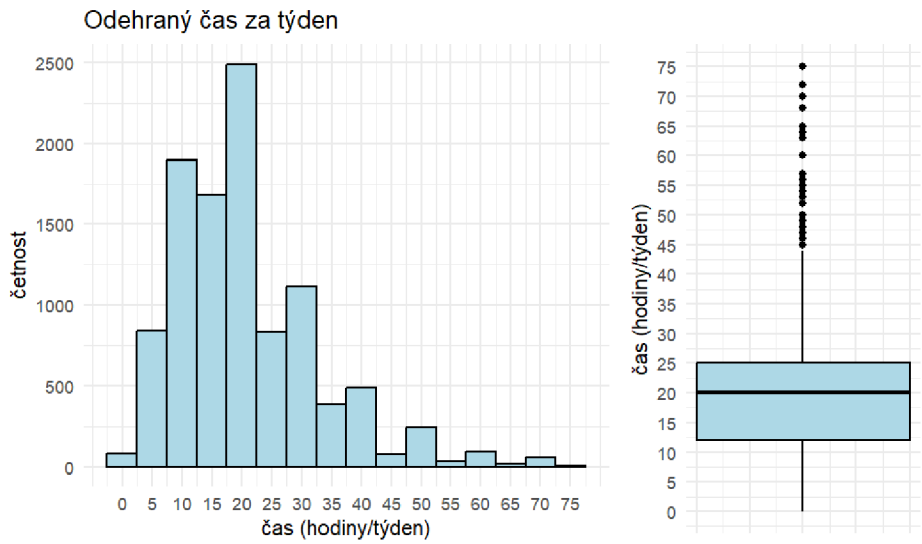
Očištěná datová sada tedy obsahuje hráče a hráčky ve věku 18-34 let pocházející z Evropy, Ameriky, Austrálie nebo Asie. Převážnou většinu však tvoří muži, lidé ve věku 18-22 let a obyvatelé Evropy a Ameriky.

5.6.5 Herní charakteristiky

Dotazník zahrnoval i otázky týkající se herních návyků. Jako statisticky významné se prokázaly následující proměnné: čas strávený hraním, čas strávený sledováním streamů, preferovaná platforma, motiv hraní a typ preferovaných spoluhráčů. Dalšími proměnnými byly konkrétní hry. Převážnou vět-

šinu tvořili hráči League of legends, na jejímž serveru byl dotazník zveřejněn. Ostatní hry byly zastoupeny v malých vzorcích, a proto se tato proměnná při modelování neuvažovala. Respondenti byli také dotázáni, zda si hraním vydělávají. Profesionálních hráčů byl jen malý vzorek, a kvůli tomu i odlišnému charakteru jejich hraní byli z dat odstraněni.

Odehraný čas byl respondenty odhadován v průměrném počtu hodin za týden. Některé odhady však byly nerealisticky vysoké, proto byl opět použit trojnásobek IQR pro očištění.

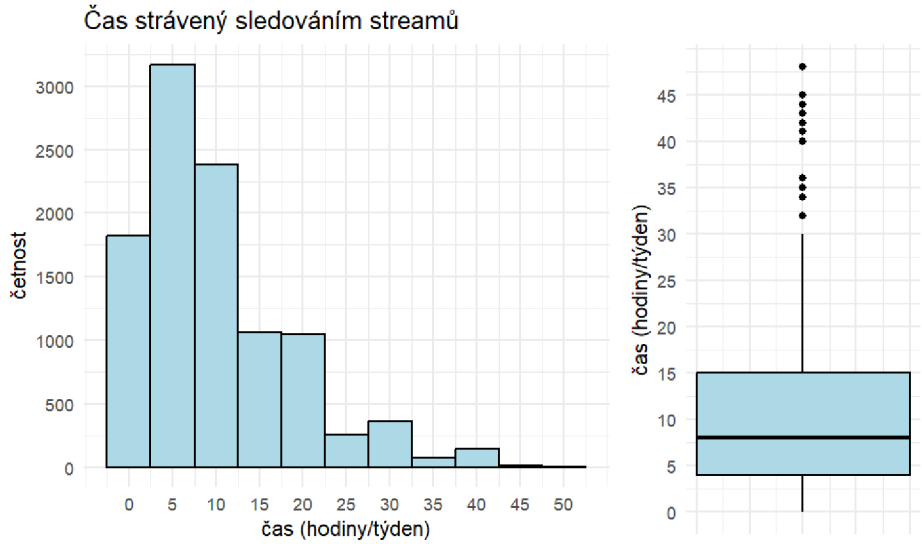


Obrázek 12: Histogram (vlevo) a krabicový graf (vpravo) respondentů podle průměrného času stráveného hraním za týden.

Po očištění v grafech na obrázku 12 vidíme, že nejvíce hráčů hraje kolem dvaceti hodin týdně. 276 lidí uvedlo, že hraje méně jak 5 hodin týdně a našli se i tací, kteří nehrají vůbec, konkrétně jich bylo 10, z nichž 4 nesledují ani streamy. 50 % respondentů hraje mezi 12-28 hodinami. Maximum, které bylo uznáno, činí 76 hodin.

Podobně hráči odhadovali počet hodin strávený sledováním streamů za týden, tedy kolik času respondenti stráví sledováním toho, jak hraje někdo

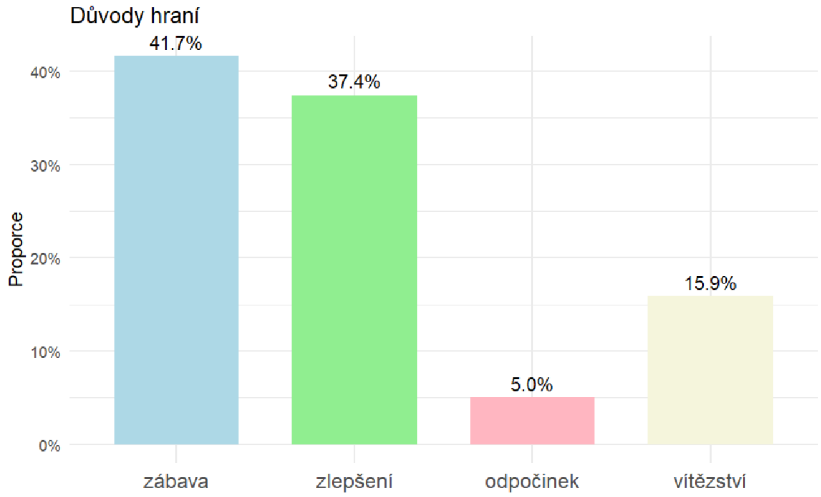
jiný. Čištění proběhlo opět pomocí trojnásobku IQR, a tak byla odstraněna extrémně vysoká pozorování.



Obrázek 13: Histogram (vlevo) a krabicový graf (vpravo) respondentů podle průměrného času stráveného sledováním streamů za týden.

Z grafů na obrázku 13 vyplývá, že 50 % respondentů sleduje streamy 4-15 hodin týdně, někteří nesledují streamy vůbec a nejvyšší uznaná hodnota činila 48 hodin.

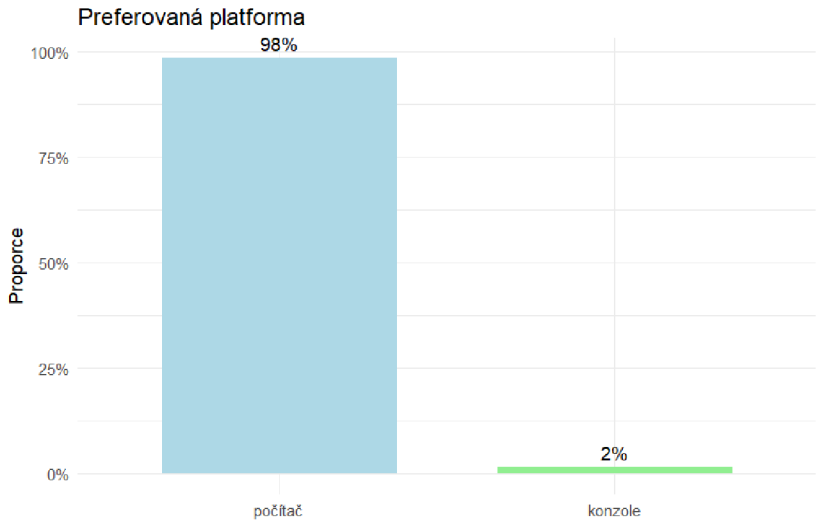
Mezi kategoriálními proměnnými byl motiv hraní, tedy hlavní důvod, proč respondent hraje hry. Respondenti měli možnost vybírat ze čtyř odpovědí, a také možnost otevřené odpovědi. Kvůli různorodosti otevřených odpovědí a složitosti jejich přeskupení byly z dat odstraněny.



Obrázek 14: Proporční četnosti motivů ke hraní her.

Z grafu na obrázku 14 je zřejmé, že převážná většina respondentů hraje hry pro zábavu nebo zlepšení svých herních dovedností. Necelých 16 % je motivováno potenciálním vítězstvím. Zbylých 5 % hledá při hraní odpočinek a odreagování.

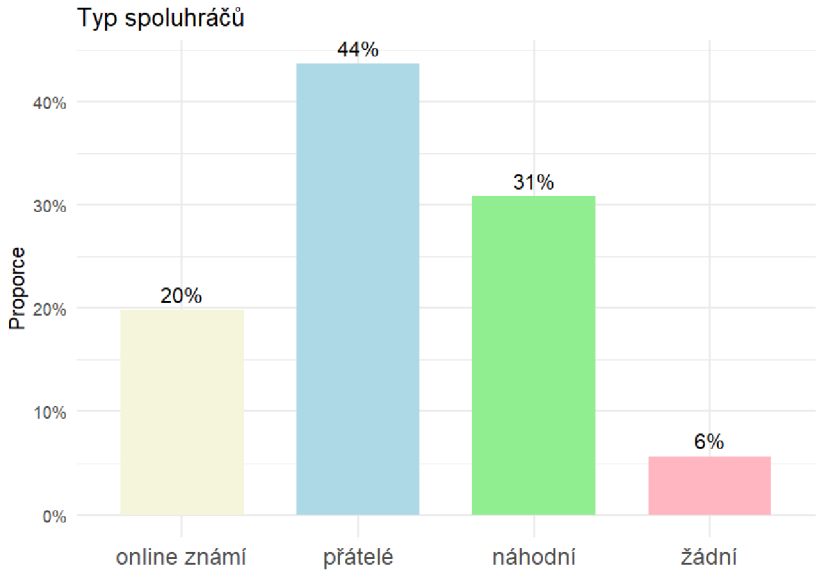
Další sledovanou proměnnou byla platforma, na které hráč hraje nejvíce. Mezi možnostmi bylo i mobilní zařízení, které však zvolilo jen pár desítek respondentů. Proto tuto platformu nebudeme v modelování uvažovat.



Obrázek 15: Proporční četnosti platforem, na kterých respondenti primárně hrají.

Z grafu na obrázku 15 je zřejmé, že převážná většina respondentů preferuje počítač. 163 hráčů upřednostňuje konzole, čímž je myšlen například Xbox nebo PlayStation. Tento úkaz bude zřejmě ovlivněn tím, že většina tázaných hráčů hraje právě League of Legends, které v roce 2017 bylo hratelné pouze na počítači. v dnešní době již existuje mobilová verze.

Následně byla položena otázka, s kým nejčastěji respondenti hrají, kde byla i možnost otevřené odpovědi. Ze stejných důvodů jako u motivů byly otevřené odpovědi odstraněny.



Obrázek 16: Proporcionální četnosti typu spoluhráčů, se kterými respondent hraje nejčastěji.

Hry se dělí primárně na dva typy: singleplayer a multiplayer. v singleplayer žánrech hráči hrají sami a zpravidla se jedná o příběhové tituly. Těmto hrám odpovídá v grafu na obrázku 16 kategorie „žádní“ a proporce činí 6 % respondentů. Zbylých 94 % procent spadá pod multiplayer hry, kde hráč hraje většinou online s jinými lidmi. Tento nepoměr je opět pravděpodobně způsoben hrou League of Legends, která je multiplayerová. Největší část hráčů hraje se svými přáteli, které znají osobně a 20 % hráčů hraje se svými známými, které poznali při hraní. Zbylých 31 % hraje s náhodně přidělenými spoluhráči, které jim hra sama najde. Často se jedná o spoluhráče na jeden zápas, tudíž k nim často nevzniká žádná vazba.

5.7 Hledání modelu

Samotné modelování probíhalo na trénovací datové sadě, která byla vytvořena stratifikovaným výběrem 90 % pozorování z původních dat. Mezi

strata patřily kategorické proměnné: pohlaví, motiv hraní, pracovní status, narcismus, spokojenost se životem, diagnózy GAD a sociální fóbie. Zbylá desetina se stala testovací sadou, na které byly schopnosti modelů ověřovány. Následně byl sestaven základní model (model0), který obsahoval proměnné: hra, platforma, typ spoluhráčů, odehraný čas, sledování streamů, věk, pohlaví, vzdělání, zaměstnanost, původ a míru narcismu.

Po dalších úvahách byly sestaveny další modely postupným přidáváním proměnných k *modelu*0. Nejprve byla přidána spokojenost se životem (dále značeno jako SWL). u modelů úzkostných poruch byla přidána pomocná otázka dotazníku GAD-7 (dále značeno jako GADE) nejprve zvlášť a poté v kombinaci se spokojeností se SWL. Následně byly tyto modely ještě vypočítány s dvojnými interakcemi. Tímto způsobem vzniklo 8 modelů pro úzkostné poruchy a 4 modely pro sociální fóbii.

Po vytvoření modelů byla použita oboustranná step funkce, která na základě informačního kritéria AIC odebírá či přidává proměnné, dokud AIC klesá. z výsledných modelů byly ještě odebrány proměnné se statisticky nevýznamnými koeficienty nebo kvůli multikolinearitě, která se objevovala v modelech s interakcemi, jelikož se stávalo, že jedna proměnná interagovala s více proměnnými. Neřešená multikolinearita může zapříčinit nespolehlivé odhady koeficientů. V modelu byla zachována ta interakce, u které byla nejsilnější statistická významnost.

Následně byla ke každému modelu vypočítala ROC křivka a na jejím základě zvolen optimální prahový bod pro klasifikaci. Klasifikování proběhlo na testovací sadě, a tak vznikla klasifikační tabulka. Nejlepší model byl zvolen na základě informací z AIC, pseudokoeficientů determinace, plochou pod ROC křivkou, přesností odhadu, senzitivitou a specificitou.

5.8 Model úzkostých poruch

Pro predikci pravděpodobnosti výskytu úzkostné poruchy bylo navrženo 8 modelů. Nyní se podíváme na jejich výkonnost.

	AIC	McFadden	Cox-Snell	Nagelkerke	AUC
základní model	8480.532	0.042	0.039	0.064	0.586
s SWL	7856.747	0.112	0.100	0.165	0.740
s GADE	6537.215	0.261	0.218	0.358	0.831
s SWL i GADE	6372.425	0.281	0.232	0.381	0.858
model s interakcemi	8467.755	0.046	0.042	0.069	0.582
int s SWL	7834.863	0.118	0.105	0.172	0.738
int s GADE	6532.178	0.263	0.219	0.359	0.830
int s SWL a GADE	6363.899	0.283	0.234	0.383	0.857

Obrázek 17: AIC, pseudokoeficienty determinace a AUC modelů úzkostí

V tabulce na obrázku 17 se ukázalo, že modely obsahující pomocnou odpověď (GADE) a spokojenost se životem (SWL) měly výrazně lepší výkonnost. Konkrétně základní model kombinující SWL a GADE dosáhl nejlepšího výsledku ve většině sledovaných metrik, včetně nejnižší hodnoty AIC (6372,425), nejvyšších pseudokoeficientů determinace McFadden (0,281), Cox-Snell (0,232) a Nagelkerke (0,381) ze všech modelů bez interakcí. Tento model předčil i většinu modelů, které interakce zahrnovaly. Jeho hodnota AUC byla také velmi vysoká (0,858), což indikuje dobrou predikční schopnost.

Jeho alternativa s interakcemi měla o něco nižší AIC (6358,853) a vyšší pseudokoeficienty determinace, avšak AUC (0,857) bylo mírně nižší než u jednoduchého modelu. i přesto je tato hodnota srovnatelná a ukazuje, že interakční model je rovněž velmi přesný.

Nyní se podívejme ještě na predikční schopnosti z klasifikační tabulky, abychom se mohli mezi posuzovanými modely lépe rozhodnout, který pro interpretaci zvolit.

	accuracy	sensitivity	specificity
základní model	0.661	0.476	0.693
s SWL	0.698	0.671	0.703
s GADE	0.798	0.713	0.812
s SWL i GADE	0.804	0.762	0.811
model s interakcemi	0.532	0.650	0.512
int s SWL	0.770	0.552	0.807
int s GADE	0.799	0.706	0.815
int s SWL a GADE	0.762	0.825	0.752

Obrázek 18: Přesnost odhadu, senzitivita a specificita modelů úzkostí

V této tabulce 18 má model s interakcemi nejvyšší senzitivitu (0,825), což je klíčové, protože naším cílem není najít klasifikační nástroj, ale popsat vztahy mezi úzkostmi a herními návyky. Proto je pro nás cennější schopnost odhalit ty, kteří úzkostmi trpí. Senzitivita modelu s interakcemi naznačuje, že je schopen dobře identifikovat jedince s úzkostnou poruchou, což je pro naše závěry zásadní. Podívejme se tedy, jak samotný model vypadá.

```
Call:
glm(formula = diagnoseGAD ~ GADE + SWL + Gender + Nar + whyplay +
    kontinent + Hours + SWL:Nar + GADE:Nar, family = binomial,
    data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4064  -0.5801  -0.2715  -0.1837   2.9660

Coefficients:
                                 Estimate Std. Error z value Pr(>|z|)
(Intercept)                     -3.735809   0.122989 -30.375  < 2e-16 ***
GADESomewhat difficult           1.905254   0.100224  19.010  < 2e-16 ***
GADEVery difficult               3.528001   0.124901  28.246  < 2e-16 ***
GADEExtremely difficult          4.456429   0.204726  21.768  < 2e-16 ***
SWLNespoko                       0.666751   0.078415   8.503  < 2e-16 ***
SWLSpok                         -0.613561   0.098045  -6.258 3.90e-10 ***
GenderFemale                     0.754262   0.114550   6.585 4.56e-11 ***
Narsilny                         0.956847   0.220747   4.335 1.46e-05 ***
whyplayimproving                -0.069224   0.075892  -0.912  0.36170
whyplayrelaxing                  0.433644   0.135826   3.193  0.00141 **
whyplaywinning                   0.287033   0.091662   3.131  0.00174 **
kontinentamerica                 0.279700   0.066350   4.216 2.49e-05 ***
kontinentaustralia               0.352794   0.217854   1.619  0.10536
kontinentasia                    0.175367   0.216898   0.809  0.41879
Hours                            0.008108   0.002600   3.119  0.00181 **
SWLNespoko:Narsilny             -0.256882   0.208537  -1.232  0.21801
SWLSpok:Narsilny                 0.440033   0.228344   1.927  0.05397 .
GADESomewhat difficult:Narsilny -0.593262   0.227175  -2.611  0.00902 **
GADEVery difficult:Narsilny     -0.716583   0.298484  -2.401  0.01636 *
GADEExtremely difficult:Narsilny -0.609100  0.499157  -1.220  0.22237
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Obrázek 19: Model úzkostných poruch

Obrázek 19 ukazuje významnost jednotlivých koeficientů, jejich odhady a statistické ukazatele pro interakční model, který zahrnuje proměnné GADE, SWL, pohlaví, narcismus, motiv hraní, kontinent, hodiny strávené hraním a jejich interakce. Pro interpretaci je třeba koeficienty převést na poměr šancí, čehož docílíme jejich vyexponováním. v dalších odstavcích se tedy budeme odkazovat na poměry šancí z tabulky na obrázku 20.

	koeficient	poměr šancí
(Intercept)	-3.73580871	0.02
GADESomewhat difficult	1.90525377	6.72
GADEVery difficult	3.52800066	34.06
GADEExtremely difficult	4.45642874	86.18
SWLNespoko	0.66675085	1.95
SWLSpok	-0.61356078	0.54
GenderFemale	0.75426176	2.13
Narsilny	0.95684679	2.60
whyplayimproving	-0.06922371	0.93
whyplayrelaxing	0.43364412	1.54
whyplaywinning	0.28703335	1.33
kontinentamerica	0.27969986	1.32
kontinentaustralia	0.35279356	1.42
kontinentasia	0.17536737	1.19
Hours	0.00810845	1.01
SWLNespoko:Narsilny	-0.25688230	0.77
SWLSpok:Narsilny	0.44003337	1.55
GADESomewhat difficult:Narsilny	-0.59326206	0.55
GADEVery difficult:Narsilny	-0.71658284	0.49
GADEExtremely difficult:Narsilny	-0.60909951	0.54

Obrázek 20: Koeficienty modelu úzkostných poruch a jejich vyexponovaná podoba

Intercept modelu zahrnuje hráče, kteří: nemají žádné problémy spojené s úzkostmi (GADE), jsou neutrálně spokojeni (SWL), jsou muži (Gender), mají slabý narcismus (Nar), hrají hry pro zábavu (whyplay) a jsou Evropané (kontinent). Pro interpretaci interceptu také předpokládáme, že spojité proměnné jsou 0. Respondent tedy průměrně nahraje 0 hodin týdně, tudíž uvažujeme, že nehraje vůbec. Takový hráč by měl podle modelu poměr šancí na přítomnost úzkostné poruchy 0,02. Je tedy velmi nepravděpodobné, že by excesivními úzkostmi trpěl.

Při dalším interpretování budeme právě s takovým typem lidí srovnávat, jak by se OR měnilo při jiných charakteristikách pozorovaného.

Podívejme se, jak se poměr šancí bude měnit právě s počtem nahraných hodin týdně. Změna na OR činí 1,01, tedy s každou nahranou hodinou vzroste o 1 %. Jak jsme viděli v krabicovém grafu (viz obrázek 12), první kvartil činil 12, medián 20 a třetí kvartil 28 hodin týdně. OR pro tyto hodnoty by bylo

1,13, 1,22 a 1,32. Nejvíce hráčů hrálo právě okolo dvaceti hodin a mají poměr šancí na úzkosti o 22 % vyšší než ti, kteří nehrají vůbec. Hráč, který hrál v datech nejvíc, tedy 76 hodin, by měl poměr šancí 2,13krát vyšší.

Můžeme se tedy na základě výsledků domnívat, že vztah mezi hraním her a úzkostmi existuje. v jiných studiích byl vztah mezi úzkostnou poruchou a závislostí na hraní popsán. Ti co generalizovanou úzkostnou poruchou trpěli, měli 8krát vyšší poměr šancí na přítomnost závislosti na hraní her. [14, 30] Naše zjištění tedy podporuje existující literaturu a naznačuje, že zvýšené hraní her může být spojeno s vyššími šancemi na úzkostné poruchy. Je důležité však zdůraznit, že není zřejmé, zda zvýšené hraní úzkosti posiluje, zda lidé s úzkostnými poruchami více hrají hry, nebo zda mají větší tendence vyplňovat podobné dotazníky, jako ten, který byl zdrojem dat.

Další herní charakteristikou byl motiv hraní, kde se neprokázal statisticky významný rozdíl mezi hráči, kteří hrají pro zábavu nebo zlepšení svých schopností. Ukázal se však u motivu výhry a odpočinku. Oproti hráčům, kteří hrají pro zábavu, bude OR u motivu výhry o 33 % vyšší a u motivu odpočinku vzroste o 54 %.

Možné vysvětlení těchto rozdílů by mohlo být následující: Hráči se během hraní setkávají samozřejmě jak s výhrou, tak prohrou. Jestliže je motivem zábava, ve hře je možné se pobavit i bez výhry. Jestliže je motivem zlepšení, může se hráč z prohry něco naučit. Pokud je ale motivem výhra samotná, může docházet v případě prohry k frustraci. Je pak pravděpodobné, že hráč bude vyhledávat další zápasy a stráví hraním více času, než původně zamýšlel. Zároveň je možné, že pro takové hráče představuje výhra jakousi formu kompenzace za neúspěchy z osobního života, což může zvyšovat predispozice k úzkostným poruchám.

Podobně jsou na tom hráči, kteří hledají v hraní relax. Hry nemusí být

vždy odpočinkové, obzvlášť když se nedaří. Hodnota koeficientu může být ovlivněna i faktory, se kterými se takoví jedinci setkávají v každodenním životě. Lidé s úzkostnými poruchami bývají často unavení, a právě zvýšená potřeba relaxace může být sama o sobě jedním z ukazatelů úzkostné poruchy.

V obou případech je tedy možné, že motiv hraní je jedním z indikátorů úzkostných poruch. Možná hry tyto úzkosti prohlubují, možná k těmto motivům lidé s úzkostmi inklinují, nebo tyto motivy zvyšují pravděpodobnost závislosti na hraní her, která je s úzkostmi spjatá. Ani jeden z těchto závěrů z analýzy přímo nevyplývá a je potřeba další výzkum, který by se blíže zaměřil na zázemí hráčů.

Ukázal se i rozdíl u pohlaví, kde ženy mají více než dvojnásobné OR oproti mužům. v datech měly ženy v průměru vyšší skór GAD-7 o dvě jednotky. Podobný efekt se uvádí i v dalších studiích, kde mají ženy obecně vyšší prevalenci úzkostných poruch než muži. To může být spojeno s mnoha faktory, jako jsou například hormonální vlivy nebo i odlišná společenská očekávání. [14][13][20] Tento rozdíl byl očekávaný a pohybuje se ve srovnatelném řádu jako v jiných výzkumech, které nebyly spojeny s hraním her.

Jediný kontinent, kde se prokázal signifikantní rozdíl od Evropských obyvatel, je Amerika, kde OR vyšlo o 32 % vyšší. Ostatní koeficienty u proměnných reprezentujících kontinenty měly sice tendenci také OR zvyšovat, ovšem statistická signifikance neobstála, což může být způsobeno nepoměrem respondentů, kteří ze zemí na daných kontinentech pocházeli.

V modelu po redukci zůstaly všechny zúčastněné proměnné pocházející z psychologických dotazníků. u kategoriální proměnné reprezentující míru komplikací způsobených úzkostmi (GADE) vyšly všechny koeficienty signifikantní. v porovnání s lidmi, kteří žádné komplikace nepociťují, mají lidé s mírnými komplikacemi téměř 7krát vyšší OR, lidé se silnými potížemi 34krát

a lidé s extrémními problémy dokonce 86krát. Tato otázka má napomáhat psychologovi k diagnostikování pacientů. Mezi hlavní kritéria v diagnostických manuálech (např. DSM-5 nebo ICD-11) patří, že dané onemocnění způsobuje v životě pacienta značné problémy. z modelu se můžeme domnívat, že taková otázka je v dotazníku skutečně směrodatná.

Proměnná SWL reprezentuje spokojenost se životem. Lidé s nižšími hodnotami SWL, tedy nespokojení, mají OR dvojnásobné. Zatímco lidé s výššími hodnotami, které bychom popsali jako velmi spokojené, mají OR poloviční. Vazba mezi spokojeností a úzkostnými poruchami je tedy zřejmá.

U silného narcismu můžeme sledovat 2,6násobný nárůst OR oproti lidem se slabou mírou narcismu. v modelu je zahrnuta i interakce SWL s narcismem, při rozšíření hladiny významnosti α alespoň na 6 %. v případě výrazné spokojenosti a silného narcismu OR roste o dalších 55 %. Spokojenost se životem však tento nárůst kompenzuje. Finální rozdíl po vynásobení všech zúčastněných OR je 2,17násobný oproti výchozí skupině. Spokojení lidé s narcismem mají sice stále dvojnásobně vyšší OR než lidé bez narcismu, nicméně oproti lidem s průměrnou mírou spokojenosti je změna v OR o 43 % nižší.

Možným vysvětlením by mohlo být, že lidé s narcismem, primárně jeho zranitelným typem, mají vyšší tendenci k úzkostem, ale skrývají je za svým sebevědomím. Takoví lidé mohou být se svou maskou v podstatě spokojeni, nebo si nechtějí nespokojenost připouštět, a přesto před úzkostmi neutečou. [22] Pro takový závěr by však musel být narcismus v datech hlouběji prozkoumán. Jedná se tedy o hypotézu pro další výzkum.

Narcismus vystupuje v modelu ještě v interakci s odpovědí GADE, kde silný narcismus signifikantně interaguje s mírnými a extrémními problémy. Výsledná změna, po vynásobení všech zúčastněných OR, je v případě mírných potíží devítinásobná a u extrémních problémů 112násobná. Je tedy

možné, že když už lidé se silnějším narcismem problémy spojené s úzkostmi zaznamenají, tak je mnohem pravděpodobnější, že se s nějakou poruchou potýkají. i zde se však jedná o hypotézu, která by mohla být předmětem dalšího zkoumání.

5.9 Model sociální fóbie

Pro predikování pravděpodobnosti sociální fóbie byly navrženy modely 4, jelikož jako jediná pomocná proměnná se nabízela spokojenost se životem (SWL). Tyto modely byly porovnávány na základě jejich výkonnosti, viz obrázek 21.

	AIC	McFadden	r2ML	r2CU	AUC
model	11382.03	0.03940361	0.04847231	0.06764029	0.6078489
s SWL	10983.35	0.07281878	0.08773229	0.12242528	0.6814790
model s interakcemi	11380.75	0.04052776	0.04982015	0.06952112	0.6098512
int s SWL	10969.19	0.07554089	0.09085826	0.12678741	0.6829733

Obrázek 21: AIC, pseudokoeficienty determinace a AUC modelů sociální fóbie

Ukázalo se, že modely obsahující SWL měly výrazně lepší výkonnost. Konkrétně model s interakcemi si vedl ve všech sledovaných metrikách nejlépe. Dosáhl nejnižší hodnoty AIC (10969,190), také měl nejvyšší hodnoty pseudokoeficientů determinace McFaddenova pseudo R^2 (0,076), Cox-Snell (0,091), Nagelkerke (0,127) a AUC (0,683), což znamená, že byl nejúspěšnější v predikci sociální fóbie. Podíváme se, jak se mu vedlo v klasifikační tabulce, viz obrázek 22

	přesnost odhadu	sensitivita	specificita
model	0.594	0.575	0.603
s SWL	0.661	0.598	0.690
model s interakcemi	0.591	0.585	0.594
int s SWL	0.645	0.621	0.655

Obrázek 22: Přesnost odhadu, senzitivita a specificita modelů sociální fóbie

Nejkomplexnější model má opět nejvyšší senzitivitu (0,621). Má tedy nejvyšší schopnost identifikovat jedince s rizikem sociální fóbie, a proto ho k popisu vztahů mezi proměnnými použijeme.

```
glm(formula = diagnoseSPIN ~ SWL + Gender + Age + streams + Hours +
    whyplay + deg + Platform + Playstyle + SWL:Hours + Gender:whyplay +
    Gender:streams + Age:Playstyle, family = binomial, data = train)

Coefficients:
                                                        Estimate Std. Error z value Pr(>|z|)
(Intercept)                                             0.869451   0.401998   2.163 0.030555 *
SWLNespoko                                              0.868114   0.112113   7.743 9.69e-15 ***
SWLSpok                                                -0.804674   0.113414  -7.095 1.29e-12 ***
GenderFemale                                            1.482934   0.162950   9.101  < 2e-16 ***
Age                                                    -0.089145   0.019182  -4.647 3.36e-06 ***
streams                                                 0.016929   0.002730   6.201 5.61e-10 ***
Hours                                                   0.008536   0.002929   2.914 0.003569 **
whyplayimproving                                       -0.137866   0.055959  -2.464 0.013751 *
whyplayrelaxing                                         0.028322   0.113721   0.249 0.803324
whyplaywinning                                          0.196634   0.069537   2.828 0.004687 **
degHigh school diploma                                 -0.137489   0.073672  -1.866 0.062009 .
degtitle                                               -0.320030   0.095178  -3.362 0.000773 ***
PlatformConsole (PS, Xbox, ...)                         0.365848   0.183793   1.991 0.046531 *
PlaystyleMultiplayer - online - with real life friends -0.443675   0.477375  -0.929 0.352679
PlaystyleMultiplayer - online - with strangers         -0.607556   0.481706  -1.261 0.207215
PlaystyleSingleplayer                                  -1.791718   0.744082  -2.408 0.016042 *
SWLNespoko:Hours                                       -0.006387   0.004356  -1.466 0.142543
SWLSpok:Hours                                           0.010304   0.004699   2.193 0.028323 *
GenderFemale:whyplayimproving                           0.035062   0.213964   0.164 0.869834
GenderFemale:whyplayrelaxing                           -1.045666   0.358369  -2.918 0.003525 **
GenderFemale:whyplaywinning                             0.077921   0.345379   0.226 0.821505
GenderFemale:streams                                   -0.022788   0.011129  -2.048 0.040587 *
Age:PlaystyleMultiplayer - online - with real life friends 0.018818 0.023078  0.815 0.414833
Age:PlaystyleMultiplayer - online - with strangers      0.028887   0.023029   1.254 0.209709
Age:PlaystyleSingleplayer                               0.095069   0.034804   2.732 0.006303 **
```

Obrázek 23: Model sociální fóbie

Na obrázku 23 vidíme, že jako statisticky signifikantní se ukázaly následující proměnné: spokojenost se životem (SWL), pohlaví (Gender), věk (Age), čas strávený na streamech (streams) a hraním her (Hours), motiv hraní (whyplay), dosažený stupeň vzdělání (deg), preferovaná platforma (platform), typ preferovaných spoluhráčů (Playstyle) a jejich interakce. Koeficienty těchto proměnných budou převedeny na poměr šancí (OR), viz obrázek 24, což umožní lepší interpretaci výsledků.

	koeficient	poměr šancí
(Intercept)	0.86945132	2.39
SWLNespoko	0.86811381	2.38
SWLSpok	-0.80467401	0.45
GenderFemale	1.48293371	4.41
Age	-0.08914487	0.91
streams	0.01692875	1.02
Hours	0.00853624	1.01
whyplayimproving	-0.13786593	0.87
whyplayrelaxing	0.02832188	1.03
whyplaywinning	0.19663365	1.22
degHigh school diploma	-0.13748854	0.87
degtitle	-0.32002981	0.73
PlatformConsole (PS, Xbox, ...)	0.36584773	1.44
PlaystyleMultiplayer - online - with real life friends	-0.44367539	0.64
PlaystyleMultiplayer - online - with strangers	-0.60755596	0.54
PlaystyleSingleplayer	-1.79171753	0.17
SWLNespoko:Hours	-0.00638718	0.99
SWLSpok:Hours	0.01030350	1.01
GenderFemale:whyplayimproving	0.03506211	1.04
GenderFemale:whyplayrelaxing	-1.04566579	0.35
GenderFemale:whyplaywinning	0.07792089	1.08
GenderFemale:streams	-0.02278850	0.98
Age:PlaystyleMultiplayer - online - with real life friends	0.01881785	1.02
Age:PlaystyleMultiplayer - online - with strangers	0.02888713	1.03
Age:PlaystyleSingleplayer	0.09506941	1.10

Obrázek 24: Model sociální fóbie

Intercept zahrnuje hráče, kteří: jsou neutrálně spokojeni, jsou muži, mají dokončené základní vzdělání, hrají na počítači, pro zábavu a s online přáteli. Za předpokladu nulových spojitých proměnných by bylo OR 0,02. Tento předpoklad by však znamenal, že mluvíme o hráčích, kteří mají 0 let.

Proto se podívejme, jak se OR bude měnit s rostoucím věkem. Jelikož proměnná věk interaguje s preferovaným typem spoluhráčů, budeme se nejprve věnovat hráčům multiplayer her, mezi kterými se neprokázal signifikantní rozdíl, ať už hrají se svými přáteli nebo s cizími lidmi. Hodnota OR pro ně činí 0,91, tedy s každým rokem života se bude o 9 % snižovat. Nejmladší respondenti měli 18 let, což znamená, že by pro ně OR kleslo přibližně o 82 %. Třetí kvartil věku respondentů na obrázku 7 představoval 22 let a pokles by činil 87 %. Nejstarší respondenti v očištěných datech měli 34 let, což by znamenalo pokles o 96 %.

Hráči singleplayer her se signifikantně lišili od hráčů multiplayerů. Když vynásobíme koeficient Age a interakční koeficient Age:PlaystyleSingleplayer, zjistíme, že se OR s každým rokem života zvyšuje o pouhé 0,1 %. Vliv věku

se tedy pro tyto hráče vyrušil a rozhodující je pouze kategoriální rozdíl, který činí pokles o 83 %. s rostoucím věkem sledujeme změny v pouhých desetinách procent, které jsou zanedbatelné. Pro hráče singleplayeru tedy neplatí, že by se OR s rostoucím věkem snižovalo.

Mezi metody kognitivně-behaviorální terapie patří expozice, která napomáhá k zvládání fóbií. Pacient je pobídnut k tomu, aby se obávaným situacím nevyhýbal nebo se jim dokonce postupně záměrně vystavoval. Může tak totiž zjistit, že scénáře, kterých se obával, se v daných situacích nenaplňují. Cílem je poučení, že situace, které úzkost vyvolávaly nejsou tak nebezpečné, jak se původně zdálo.

Vlivem socializace a zkušeností, které s rostoucím věkem nabýváme, se zřejmě riziko sociální fóbie snižuje nebo se s ní učíme vypořádat. o poklesu prevalence sociální fóbie s věkem psali i jiní autoři. [18]

Hraní online her určitou formu socializace představuje, jelikož je často žádoucí, aby komunikace a vzájemná spolupráce mezi spoluhráči probíhala. Je tedy možné, že i hraní online her posiluje komunikační schopnosti a tím snižuje úzkosti spojené se socializací. Také je však možné, že tyto hry spíše preferují jedinci, kteří s komunikací problémy nemají a socializace je jim příjemná. Jednoznačný závěr ovšem z modelu není zřejmý. u hráčů singleplayer her se však tento trend neprokázal. Možná tyto hry právě přitahují hráče se sociální fóbií, jelikož hrají sami a vyhýbají se kontaktu s ostatními lidmi i v online prostoru. Také je možné, že tito hráči obecně preferují méně sociálního kontaktu, tudíž vlivem izolace, kterou i singleplayer hry představují, se problémy spojené se sociální fóbií prohlubují nebo se aspoň s věkem nezlepšují. Nevýhodou modelů je, že vyšší věkové kategorie zastupoval jen malý poměr respondentů, proto také mohlo dojít ke zkreslení výsledků.

Dalšími signifikantními proměnnými byl nejvyšší dosažený stupeň vzdě-

lání a preferovaná platforma. Uvažujeme-li hladinu významnosti alespoň 7%, považujeme koeficient u maturitního vzdělání za významný. Lidé s tímto stupněm vzdělání budou mít OR o 13 % nižší než lidé se základním. Ti, co absolvovali vysokou školu, mají OR o 27 % nižší. Co se týče herní platformy, hráči na konzolích mají OR o 44 % vyšší než hráči na počítačích.

Zdá se, že čím vyššího vzdělání dosáhneme, tím menší je riziko sociální fóbie. Vzdělání souvisí s věkem, což může být jeden z důvodů, proč v modelu pozorujeme pokles. Možná je tento pokles způsoben i tím, že s vyšším vzděláním přibývá zkušeností, které mají i sociální charakter. Příkladem mohou být ústní zkoušky, kdy se student musí naučit obhájit své znalosti, což posiluje jeho komunikační dovednosti a v případě úspěchu i sebevědomí. Zároveň může roli hrát i status, který vyšší vzdělání přináší, což sebevědomí dále posiluje. Některé studie přímo spojují sociální fóbii s nízkým stupněm vzdělání [8], zatímco jiné tvrdí, že sociální fóbie je jedním z prediktorů toho, že studenti vyšší vzdělání nedokončí [24].

Preferovaná platforma souvisí s hraním offline her. Pozorovaný rozdíl může být částečně vysvětlen tím, že hráči na konzolích tráví více času hraním singleplayer her, které jsme již dříve identifikovali jako potenciálně zvyšující riziko sociální fóbie. Online hry se obecně lépe hrají s klávesnicí a myší, zatímco singleplayer hry s ovladačem, kterým konzole automaticky disponují. Navíc u většiny konzolí je pro online hraní potřeba dalšího předplatného, které si nepořídí každý. Zřejmě jde o další prohloubení vztahu mezi sociální fóbií a singleplayer hrami. v datech však hráči na konzolích představovali jen 2 %, a proto je možné zkreslení výsledků.

Spokojenost se životem se v modelu projevila v interakci s odehraným časem. u samotné proměnné se v obou kategoriích ukázaly rozdíly jako statisticky významné. Jestliže je dotyčný nadmíru nespokojen, OR je 2,4krát

vyšší ve srovnání s neutrálně spokojenými. Pro neutrálně spokojené hráče se OR zvýší o 1 % s každou nahranou hodinou týdně. První kvartil počtu nahraných hodin činil 12 a nárůst na OR by představoval 12 %. Medián byl 20 s nárůstem o 22 % a třetí kvartil 28 s nárůstem o 32 % oproti nehrajícím lidem. Nejvyšší uznaná hodnota činila 76 hodin a dle modelu by byl nárůst OR dvojnásobný. Pro nespokojené hráče je vliv odehraného času stejný.

V interakci s časem se však projevil významný rozdíl u hráčů, kteří jsou nadmíru spokojení. Oproti neutrálně spokojeným lidem mají OR o 55 % nižší, ovšem s každou odehranou hodinou týdně stoupá o 2 %. Pokud spokojení hráči hrají 12 hodin týdně, tak OR klesne o 43 %, což je o něco méně než u spokojených lidí, kteří nehrají vůbec. Hrají-li 28 hodin týdně, potom OR klesne o 22 %. Zlomový bod je ve 41 hodinách, kdy se pokles spojený se spokojeností překoná a klesající charakter se změní na rostoucí. v případě nejvyšší hodnoty 76 je nárůst OR opět dvojnásobný. Spokojenost se životem tedy OR snižuje, ovšem v kombinaci s hraním her pouze do 40 hodin týdně. Potom začne OR nabývat podobných hodnot, jako u neutrálně spokojených hráčů.

V modelu pozorujeme podobný vztah s hraním, jako byl u úzkostných poruch. Sociální fóbie se řadí mezi úzkostné poruchy a stejně tak i souvisí se závislostí na hraní her. Podle některých studií se hráči se sociální fóbií mohou silněji identifikovat s herními postavami (roleplay), čímž se prohlubují symptomy závislosti na hrání. [24] Hry zahrnující roleplay jsou často singleplayerové. Je možné, že si tak hráči suplují neuspokojivé sociální interakce z reálného života. Hry se tak stávají jejich bezpečným prostředím, ve kterém se mohou svobodně projevovat, a proto v nich chtějí trávit více času i za cenu izolace od reálného světa. Skuteční lidé však málokdy interagují stejně jako herní postavy, což možná vede k prohlubování sociální fóbie

stejně jako závislosti na hraní.

Tato teorie by mohla také vysvětlovat rostoucí OR při spokojenosti se životem a hraním nad 40 hodin týdně (tj. téměř 6 hodin denně). Hráči mohou být se svým životem ve hře spokojeni, i když se primárně realizují ve hře. Taková situace však může vést k většímu množství času stráveného ve virtuálním světě a k další izolaci od světa reálného, což by mohlo dále zhoršovat symptomy sociální fóbie a závislosti na hraní.

Jiné studie však tvrdí, že jsou-li roleplay hry skupinovou aktivitou, kdy hrajeme s přáteli, může se intenzita symptomů sociální fóbie a úzkostí snižovat. Tato aktivita je také navrhována jako intervence pro pacienty trpící sociální fóbií. [28] Zřejmě je tak naplněna potřeba socializace z bezpečí domova. Nevýhodou dat, na kterých byl náš model sociální fóbie trénován, je nízký poměr respondentů hrajících roleplay hry, protože League of Legends mezi ně nepatří. Jedná se tak o zajímavé hypotézy pro další výzkum.

V případě sledování streamů se OR zvyšuje o 2 % s každou shlédnutou hodinou týdně. Růst je tedy vyšší, jak u hraní her, ovšem počet hodin strávený sledováním byl v datech výrazně nižší oproti hraní. První kvartil na boxplotu (viz Obr. 13) představoval 4 hodiny, což znamená nárůst o 8 %. Medián činil 8 s nárůstem o 17 %. Třetí kvartil 15 s nárůstem o 34 %. Hráči sledující 35 hodin týdně už mají nárůst dvojnásobný oproti těm, co nesledují vůbec.

Sledování streamů se prokázalo jako statisticky významné v interakci s pohlavím, tudíž předešlé změny na OR platí pouze pro muže. Ženy mají OR v porovnání s muži 4,4násobné. Tento rozdíl byl očekávaný, jelikož podle některých zdrojů jsou ženy náchylnější k sociální fóbii a jsou více ochotné si ji přiznat. [8][18] Ovšem v interakci se sledováním streamů OR naopak o 2 % klesá s každou hodinou. Žena, která sleduje streamy 15 hodin týdně má

nárust 3,25. Je sice mnohem výraznější než u mužů, ale nižší než u hráček, které streamy nesledují vůbec.

Sledování streamů se ukázalo jako významnější ukazatel, jelikož u mužů stačí téměř polovina hodin než u hraní her, pro výrazný nárůst na poměru šancí výskytu sociální fóbie. Možným vysvětlením by mohlo být, že sledování streamů je poměrně pasivní činnost, u které nemusí docházet k téměř žádnému sociálnímu kontaktu. Proto je možná sledování streamů příjemnější pro muže, kteří mají rádi hry, ale potýkají se s problémy spojenými se sociální fóbií. k socializaci však při něm nedochází, a tak se mohou symptomy fóbie prohlubovat.

Paradoxně u hráček sledování streamů OR snižuje. Pro ženy bývají online hry poměrně toxickým prostředí a můžeme se na internetu setkat se spoustou videí, kde jsou ženy během hraní ponižovány nebo dokonce obtěžovány, což může sociální fóbii prohlubovat. Jedním z možných vysvětlení, proč sledování streamů OR u žen snižuje je, že sledování streamů poskytuje formu bezpečné náležitosti a zapojení do herní komunity bez přímé interakce, která by mohla být stresující nebo toxická. Tímto způsobem je možné, že u žen sledování streamů působí jako ochranný faktor proti sociální fóbií. Opět se však jedná pouze o hypotézy pro další výzkum.

Poslední proměnnou, která se prokázala statisticky signifikantní je motiv hraní. Tato proměnná se také ukázala jako významná v interakci s pohlavím. Muži, kteří hrají pro zlepšení svých schopností, mají OR o 13 % nižší než hráči hrající pro zábavu. Oproti tomu hráči hrající pro výhru mají OR o 22 % vyšší. Hráčky, které hrají pro relaxaci mají OR o 65 % nižší než jiné hráčky.

Lidé se sociální fóbií často trpí silnými obavami z hodnocení druhými. Možná právě to odhaluje motiv výhry, jelikož výhra a celkový úspěch ve hře podvědomě slibuje chválu a úctu spoluhráčů. Tento motiv může také zakrývat

strach z prohry, kdy by naopak mohlo dojít ke kritice herního výkonu a případným útokům na osobu hráče. Tato motivace má tedy možná zčásti vnější charakter, kdy se hráči snaží vyhnout negativnímu hodnocení spoluhráčů, či výsměchu oponentů. Motivace zlepšení však může spíše představovat vnitřní zadostiučinění, kdy hráč vnímá svůj progres bez nutné závislosti na hodnocení okolím. Motivace zlepšování je možná i efektivnější strategií ke zvýšení šancí na výhru. Proto se i takoví hráči zřejmě spíše setkají s pozitivním hodnocením díky svým vypilovaným schopnostem.

Většina respondentů v datech pocházela ze hry League of Legends, kde je často možné sledovat hráče, kterým se nedaří, jak aktivně viní z blížící se prohry své spoluhráče nebo dokonce celý tým. Je možné, že někteří z těchto hráčů mají takové obavy z kritiky, že se jí snaží předejít shozením viny na někoho jiného a upozorňováním na jejich nedostatky při hraní. Toto chování může plynout ze sociální fóbie. Paradoxně je strategie obviňování z většiny případů nefunkční a často vede k spolčení týmu proti urážejícímu spoluhráči.

V případě hráček je možné, že ty, co hrají pro relax, jsou vůči hodnocení okolím z části imunní a nenechají se jím rozhodit. Jak už bylo zmíněno dříve, hráčky to mívají v převážně mužské komunitě hráčů těžší. Jestli jim však hraní relax poskytuje, tak je možné, že jsou vyrovnanější a komentářům od spoluhráčů se nemusí obávat. Možná proto je u nich přítomnost sociální fóbie méně pravděpodobná.

Závěr

Cílem této práce bylo popsat vztahy mezi herními návyky a úzkostnými poruchami pomocí koeficientů logistické regrese. v teoretické části byl odvozen tvar modelu logistické regrese a představena interpretace koeficientů pomocí poměru šancí, to vše v souvislosti s lineární regresí pro intuitivní pochopení. v druhé kapitole byla představena metoda maximální věrohodnosti, jak z obecného hlediska, tak její aplikace v kontextu nalezení koeficientů logistické regrese. Třetí kapitola byla věnována kritériím pro nalezení vhodného modelu pomocí AIC, pseudokoeficientů determinace a Waldova testu, který ověřuje významnost koeficientů modelu. Ve čtvrté kapitole jsme se zaměřili na hodnocení binárního klasifikátoru prostřednictvím senzitivity, specificity a ROC křivek. Cílem bylo tyto složité koncepty vysvětlit co nejjednodušeji a nejlépe na příkladech pro posílení intuice.

V praktické části jsme si představili datovou sadu, ze které jsme při analýze čerpali. Zároveň byly stručně představeny psychologické proměnné, společně s psychometrickými inventáři, které byly použity pro sběr dat.

Analýza ukázala, že zvýšené riziko úzkostné poruchy je spojeno s časem stráveným hraním her. Toto zjištění podporuje teorii, že závislost na hraní může úzkosti u hráčů posilovat. Významnou roli hraje také motivace, se kterou hráči hry hrají. Pro úzkostnou poruchu se jako nejrizikovější ukázal motiv relaxace, následovaný motivem výhry, který se také ukázal jako nejrizikovější pro zvýšené šance sociální fóbie. Pohlaví se rovněž ukázalo jako velmi významný faktor, kdy ženy mají několikanásobně vyšší šance na úzkostné poruchy než muži.

V obecném modelu úzkostných poruch se jako nejvýznamnější faktory zvýšení šance ukázaly narcismus a nespokojenost se životem. Významně vyšší rozdíly byly také zjištěny u hráčů pocházejících z Ameriky oproti Evropě.

Otázkou zůstává, zda tyto rozdíly zaznamenáváme obecně v populaci, nebo jak se liší v případě, že dané osoby jsou hráči počítačových her.

Z modelu sociální fóbie lze usuzovat, že online hry mohou představovat určitou formu socializace, která by v některých případech mohla fóbií zmírňovat. Klíčová je motivace a čas strávený hraním, jelikož pozitivní efekt by mohl být překonán závislostí na hraní. Problematičtější jsou hry, kde hráči hrají sami, nebo sledování streamů, které by mohly vlivem absence sociálního kontaktu fóbií posilovat. Analýza také podporuje teorii, že s věkem a vzděláním riziko sociální fóbie klesá. Není však jasné, do jaké míry jsou tato zjištění přirozená pro obecnou populaci a jak se liší v případě hráčů.

Nevýhodou dotazníkových dat je, že nemůžeme s jistotou říci, jak dobře reprezentují skutečnost. Závisí totiž na tom, jací hráči se rozhodli dotazník vyplnit, jak byli při vyplňování poctiví a jakou formou byl dotazník podáván. Data obsahovala výrazné nepoměry určitých skupin, jako jsou ženy, hráči na jiných platformách, než je počítač, hráči z jiných kontinentů než Evropa a Amerika, věkové rozmezí nebo různorodost her. Zároveň není vhodné diagnostikovat poruchy čistě na základě jednoho vyplnění dotazníku. Například dotazník GAD-7 se ptá na úzkosti v posledních dvou týdnech, což může znamenat, že diagnostikovaný respondent může procházet jen dočasnou tíživou situací, ale trvalou úzkostnou poruchou netrpí. Tyto faktory mohly analýzu zkreslit. Je tudíž zásadní nahlížet na výsledné interpretace spíše jako na hypotézy pro další výzkum a nepovažovat je za finální závěry.

Pro mě jako autora byla tato práce velmi přínosná především jako praxe, kde jsem si mohl vyzkoušet komplexní datovou analýzu. Uvědomil jsem si, jak složité je interpretovat výsledky analýz a jak křehká mohou být výsledná tvrzení. Zároveň jsem měl možnost lépe pochopit metodu lineární regrese především z praktického hlediska a rozšířil jsem si své obzory o tom, jak se

píše odborný text. Tyto zkušenosti jsou pro mě nedocenitelné a považuji je za vrchol svého bakalářského studia. Jsem za ně velmi vděčný.

Literatura

[1] Agresti, A.: Categorical Data Analysis. Hoboken, New Jersey: John Wiley & Sons, Inc., 2002.

[2] Agresti, A.: Foundations of Linear and Generalized Linear Models. Hoboken, New Jersey: John Wiley & Sons, Inc., 2015.

[3] American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*

[4] Bortlíček, Z.: ROC křivky. Brno, 2008. Diplomová práce. Masarykova univerzita v Brně, Přírodovědecká fakulta. Dostupné z: https://is.muni.cz/th/jwg4n/Diplomovaprace.pdf.

[5] Crocq, M. A. (2017). The history of generalized anxiety disorder as a diagnostic category. Dialogues in clinical neuroscience, 19(2), 107-116.

[6] Diener, E., Emmons, R. A., Larsen, R. J. & Griffin, S. (1985) The Satisfaction With Life Scale, Journal of Personality Assessment, 49:1, 71-75, DOI: 10.1207/s15327752jpa4901_13

[7] Festl, R., Scharkow, M., & Quandt, T. (2013). Problematic computer game use among adolescents, younger and older adults. Addiction, 108(3), 592-599.

[8] Furmark, T., Tillfors, M., Everz, P. O., Marteinsdottir, I., Gefvert, O., & Fredrikson, M. (1999). Social phobia in the general population: prevalence and sociodemographic profile. Social psychiatry and psychiatric epidemiology, 34, 416-424.

[9] https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/faq-what-are-pseudo-r-squareds/

[10] Hron, K.; Kunderová, P. a Vencálek, O., Základy počtu pravděpodobnosti a metod matematické statistiky. 3. přepracované vydání. Skripta. Olomouc: Univerzita Palackého v Olomouci, 2018. ISBN 978-80-244-5398-9.

[11] Jefferies, P., & Ungar, M. (2020). Social anxiety in young people: a prevalence study in seven countries. PloS one, 15(9), e0239133.

[12] Konrath, Sara, Brian P. Meier, and Brad J. Bushman. "Development and validation of the single item narcissism scale (SINS)."PLOS one 9.8 (2014): e103469.

[13] McLean, C. P., Asnaani, A., Litz, B. T., & Hofmann, S. G. (2011). Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. Journal of psychiatric research, 45(8), 1027-1035.

[14] Huang, X., Shi, H. X., Li, H. Q., Guo, W. J., Luo, D., & Xu, J. J. (2022). Differential effects of anxiety on Internet gaming disorder: a large-scale cross-sectional survey. Frontiers in Psychiatry, 12, 802513.

[15] Mentzoni, R. A., Brunborg, G. S., Molde, H., Myrseth, H., Skouverøe, K. J. M., Hetland, J., & Pallesen, S. (2011). *Problematic video game use: estimated prevalence and associations with mental and physical health. Cyberpsychology, behavior, and social networking*, 14(10), 591-596.

[16] Nerys Williams, The GAD-7 questionnaire, Occupational Medicine, Volume 64, Issue 3, April 2014, Page 224, https://doi.org/10.1093/occmed/kqt161

[17] Niermann, H. C., Voss, C., Pieper, L., Venz, J., Ollmann, T. M., & Beesdo-Baum, K. (2021). Anxiety disorders among adolescents and

young adults: Prevalence and mental health care service utilization in a regional epidemiological study in Germany. Journal of Anxiety Disorders, 83, 102453.

[18] Ohayon, M. M., & Schatzberg, A. F. (2010). Social phobia and depression: prevalence and comorbidity. Journal of psychosomatic research, 68(3), 235-243.

[19] Pecáková, I. (2007). Logistická regrese s vícekategoriální vysvětlovanou proměnnou. Acta Oeconomica Pragensia, 1, 86-86.

[20] Pesce L, van Veen T, Carlier I, et al. Gender differences in outpatients with anxiety disorders: the Leiden Routine Outcome Monitoring Study. Epidemiology and Psychiatric Sciences. 2016;25(3):278-287. doi:10.1017/S2045796015000414

[21] Ranta, K., Kaltiala-Heino, R., Rantanen, P., Tuomisto, M., & Marttunen, M. (2007). Screening social phobia in adolescents from general population: The validity of the Social Phobia Inventory (SPIN) against a clinical interview. European Psychiatry, 22(4), 244-251. doi:10.1016/j.eurpsy.2006.12.002.

[22] Röhr, H.P., Narcismus - vnitřní žalář: vznik poruchy, průběh a možnosti jejího překonání. Vydání třetí. Přeložil Petr PATOČKA. Praha: Portál, 2016. Spektrum (Portál)

[23] Sauter, M., & Draschkow, D. (2017). Are Gamers Sad and Isolated? a database about the Anxiety, Life Satisfaction and Social Phobia of over 13000 participants.

[24] Sioni, S. R., Burleson, M. H., & Bekerian, D. A. (2017). Internet gaming

disorder: Social phobia and identifying with your virtual self. Computers in Human Behavior, 71, 11-15.

[25] Stein, M. B., & Stein, D. J. (2008). Social anxiety disorder. The lancet, 371(9618), 1115-1125.

[26] Useche, S., & Serge, A. (2016). The satisfaction with life scale (SWLS): Psychometric properties and observed scores in university students. mental health, 14, 34.

[27] Urbánek, T., DENGLEROVÁ D. a Širůček J.. Psychometrika: Měření v psychologii. Praha: Portál, 2011, 320 s. ISBN 978-80-7367-836-4.

[28] Varrette, M., Berkenstock, J., Greenwood-Ericksen, A., Ortega, A., Michaels, F., Pietrobon, V., & Schodorf, M. (2022). Exploring the efficacy of cognitive behavioral therapy and role-playing games as an intervention for adults with social anxiety. Social Work With Groups, 46(2), 140–156. https://doi.org/10.1080/01609513.2022.2146029

[29] Vykydalová, V.. Metoda maximální věrohodnosti a podobné techniky odhadu parametrů. Olomouc, 2021. diplomová práce (Mgr.). UNIVERZITA PALACKÉHO v OLOMOUCI. Přírodovědecká fakulta

[30] Wang, C. Y., Wu, Y. C., Su, C. H., Lin, P. C., Ko, C. H., & Yen, J. Y. (2017). Association between Internet gaming disorder and generalized anxiety disorder. Journal of behavioral addictions, 6(4), 564-571.

[31] Weinstein, A. M. (2010). *Computer and Video Game Addiction—A Comparison between Game Users and Non-Game Users.* The American Journal of Drug and Alcohol Abuse, 36(5), 268–276. https://doi.org/10.3109/00952990.2010.491879

[32] World Health Organization (2018). *International Classification of Diseases, Eleventh Revision (ICD-11)*

[33] Zamani, E., Kheradmand, A., Cheshmi, M., Abedi, A., & Hedayati, N. (2010). Comparing the social skills of students addicted to computer games with normal students. Addiction & health, 2(3-4), 59.

[34] Zlámal, F.. *Logistická regrese*. Online. Bachelor's thesis. Brno: Masaryk University, Faculty of Science. 2013. Available from: https://theses.cz/id/2gui9h/.

Přílohy

Příloha č. 1 Zdroj dat a dotazník

Data a dotazník, ze kterého tato práce vychází, můžete najít na odkazu: https://osf.io/vnbxk/

Příloha č. 2 Dotazník GAD-7

Over the last 2 weeks, how often have you been bothered by the following problems? *

	Not at all	Several days	Over half the days	Nearly every day
1. Feeling nervous, anxious, or on edge	○	○	○	○
2. Not being able to stop or control worrying	○	○	○	○
3. Worrying too much about different things	○	○	○	○
4. Trouble relaxing	○	○	○	○
5. Being so restless that it's hard to sit still	○	○	○	○
6. Becoming easily annoyed or irritable	○	○	○	○
7. Feeling afraid as if something awful might happen	○	○	○	○

If you checked off any problems, how difficult have these made it for you to do your work, take care of things at home, or get along with other people?

- ○ Not difficult at all
- ○ Somewhat difficult
- ○ Very difficult
- ○ Extremely difficult

Příloha č. 3 Dotazník SPIN

	Not At All	A Little Bit	Somewhat	Very Much	Extremely
1. I am afraid of people in authority.	○	○	○	○	○
2. I am bothered by blushing in front of people.	○	○	○	○	○
3. Parties and social events scare me.	○	○	○	○	○
4. I avoid talking to people I don't know.	○	○	○	○	○
5. Being criticized scares me a lot.	○	○	○	○	○
6. I avoid doing things or speaking to people for fear of embarrassment.	○	○	○	○	○
7. Sweating in front of people causes me distress.	○	○	○	○	○
8. I avoid going to parties.	○	○	○	○	○
9. I avoid activities in which I am the center of attention.	○	○	○	○	○
10. Talking to strangers scares me.	○	○	○	○	○
11. I avoid having to give speeches.	○	○	○	○	○
12. I would do anything to avoid being criticized.	○	○	○	○	○
13. Heart palpitations bother me when I am around people.	○	○	○	○	○
14. I am afraid of doing things when people might be watching.	○	○	○	○	○
15. Being embarrassed or looking stupid are among my worst fears.	○	○	○	○	○
16. I avoid speaking to anyone in authority.	○	○	○	○	○
17. Trembling or shaking in front of others is distressing to me.	○	○	○	○	○

Příloha č. 4 Dotazník SINS

To what extent do you agree with this statement: 'I am a narcissist.

Note: The word 'narcissist' means egotistical, self-focused and vain.

- ◯ 1 - not very true of me
- ◯ 2 - somewhat true of me
- ◯ 3 - moderately true of me
- ◯ 4 - mostly true of me
- ◯ 5 - very true of me

Příloha č. 5 Dotazník SWLS

2. Questionnaire B

DIRECTIONS: Below are five statements with which you may agree or disagree. Using the 1 - 7 scale below, indicate your agreement with each item by crossing the appropriate item. Please be open and honest in your responding.

*

	Strongly Disagree	Disagree	Slightly Disagree	Neither Agree or Disagree	Slightly Agree	Agree	Strongly Agree
1. In most ways my life is close to my ideal.	○	○	○	○	○	○	○
2. The conditions of my life are excellent.	○	○	○	○	○	○	○
3. I am satisfied with life.	○	○	○	○	○	○	○
4. So far I have gotten the important things I want in life.	○	○	○	○	○	○	○
5. If I could live my life over, I would change almost nothing.	○	○	○	○	○	○	○