

Efekt domina w edukacji: Jak środowisko rodzinne i społeczne kształtuje zachowania, wyniki i aspiracje uczniów

Maksym Dunajewski, Idzi Łopatniuk

2025-06-04

Wprowadzenie

Uczniowie nie funkcjonują w próżni – ich zachowania, wybory edukacyjne i życiowe są silnie kształtowane przez otaczające ich środowisko. Relacje rodzinne, status społeczny, jakość wsparcia edukacyjnego czy styl życia to czynniki, które wzajemnie na siebie oddziałują, tworząc złożony system powiązań.

Niektóre z tych zależności mogą działać jak efekt domina – pojedynczy element, taki jak sytuacja rodzinna, może uruchomić ciąg powiązanych zmian wpływających na zaangażowanie ucznia, jego wyniki w nauce oraz decyzje dotyczące przyszłości.

W ramach tego projektu przyjmujemy za główne założenie, że **środowisko rodzinne i społeczne ucznia wywiera efekt domina na jego styl życia, wyniki edukacyjne oraz aspiracje dotyczące dalszej edukacji.**

Opis zbioru danych

Dane wykorzystane w projekcie pochodzą z anonimowego badania przeprowadzonego wśród uczniów portugalskich szkół średnich, uczęszczających na kursy z matematyki oraz języka portugalskiego. Zbiór zawiera szeroki zakres informacji socjodemograficznych, edukacyjnych oraz behawioralnych, umożliwiających kompleksową analizę czynników wpływających na życie i osiągnięcia uczniów.

Dostępne zmienne obejmują m.in. dane dotyczące:

- szkoły, płci, wieku i miejsca zamieszkania ucznia,
- struktury i statusu rodziny, poziomu wykształcenia rodziców oraz ich zawodów,
- relacji rodzinnych i opiekuńczych, dostępu do Internetu, statusu związku,
- nawyków i zachowań (np. czas nauki, spożycie alkoholu, aktywność towarzyska, udział w zajęciach dodatkowych),
- przebiegu edukacji (liczba niepowodzeń, obecność na lekcjach, udział w korepetycjach),
- oraz ocen uzyskanych w trzech okresach nauki (G1, G2, G3).

Chociaż pierwotnym celem autorów danych było przewidywanie wyników końcowych uczniów, ich struktura pozwala również na prowadzenie szerszych analiz statystycznych i społecznych. Zróżnicowanie zmiennych sprawia, że zbiór stanowi doskonałe źródło do badania złożonych zależności między środowiskiem życia ucznia, jego stylem funkcjonowania a decyzjami edukacyjnymi.

W projekcie skoncentrowano się na analizie ciągów zależności przyczynowo-skutkowych zawartych w danych, które doskonale odzwierciedlają założenia hipotezy o „efekcie domina” w środowisku szkolnym i społecznym uczniów.

Źródło danych:

P. Cortez and A. Silva (2008), *UCI Machine Learning Repository – Student Alcohol Consumption*,

<https://archive.ics.uci.edu/ml/datasets/STUDENT+ALCOHOL+CONSUMPTION>

Eksploracja Danych

```
library(tidyverse)
library(ggplot2)
library(dplyr)
library(scales)
library(ResourceSelection)
library(caret)
library(showtext)

d1=read.table("student-mat.csv",sep=",",header = TRUE)
d2=read.table("student-por.csv",sep=",",header = TRUE)
students <- rbind(d1, d2)

students <- students %>%
  mutate(
    famrel_good = ifelse(famrel >= 3, 1, 0),
    high_alcohol = ifelse(Walc >= 4 | Dalc >= 4, 1, 0),
    ideal_grade = ifelse(G3>= 15, 1, 0)
  )

summary(students %>% select(G3, famrel, Walc, Dalc, goout, freetime, studytime, higher))

##          G3           famrel          Walc          Dalc
##  Min.   : 0.00   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:10.00  1st Qu.:4.000   1st Qu.:1.000   1st Qu.:1.000
##  Median :11.00  Median :4.000   Median :2.000   Median :1.000
##  Mean   :11.34  Mean   :3.936   Mean   :2.284   Mean   :1.494
##  3rd Qu.:14.00  3rd Qu.:5.000   3rd Qu.:3.000   3rd Qu.:2.000
##  Max.   :20.00  Max.   :5.000   Max.   :5.000   Max.   :5.000
##          goout          freetime         studytime        higher
##  Min.   :1.000   Min.   :1.000   Min.   :1.00   Length:1044
##  1st Qu.:2.000   1st Qu.:3.000   1st Qu.:1.00   Class :character
##  Median :3.000   Median :3.000   Median :2.00   Mode  :character
##  Mean   :3.156   Mean   :3.201   Mean   :1.97
##  3rd Qu.:4.000   3rd Qu.:4.000   3rd Qu.:2.00
##  Max.   :5.000   Max.   :5.000   Max.   :4.00
```

Etap 1: Wpływ środowiska rodzinnego na relacje w rodzinie

Pierwszym ogniwem analizowanego efektu domina jest środowisko rodzinne ucznia, rozumiane przez pryzmat wykształcenia rodziców, struktury rodziny oraz sytuacji opiekuńczej. Zakładamy, że czynniki te wpływają na jakość relacji w rodzinie, która może mieć kluczowe znaczenie dla dalszego funkcjonowania ucznia.

W tej części analizie poddano zależność pomiędzy zmiennymi opisującymi środowisko rodzinne (wykształcenie matki i ojca, status rodziców, liczebność rodziny) a oceną relacji rodzinnych (famrel). W celu uproszczenia analizy, zmienną famrel zaklasyfikowano binarnie: jako dobre relacje (famrel ≥ 3) oraz negatywne jako pozostałe przypadki.

Celem tego etapu jest sprawdzenie, czy lepsze warunki rodzinne wiążą się z większym prawdopodobieństwem pozytywnych relacji w domu, co stanowi fundament dalszych elementów analizowanego ciągu przyczynowo-skutkowego.

```
# Model: wpływ środowiska rodzinnego na relacje
model_famrel <- glm(famrel_good ~ Medu + Fedu + Pstatus + famsize,
                      data = students, family = binomial)

# Podsumowanie wyników
summary(model_famrel)
```

```
##
## Call:
## glm(formula = famrel_good ~ Medu + Fedu + Pstatus + famsize,
##       family = binomial, data = students)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 2.18678   0.45585  4.797 1.61e-06 ***
## Medu        -0.09180   0.13846 -0.663  0.5073
## Fedu         0.02047   0.13838  0.148  0.8824
## PstatusT     0.69780   0.31056  2.247  0.0246 *
## famsizeLE3   -0.16544   0.25863 -0.640  0.5224
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 549.66 on 1043 degrees of freedom
## Residual deviance: 542.91 on 1039 degrees of freedom
## AIC: 552.91
##
## Number of Fisher Scoring iterations: 5
```

Wyniki modelu: wpływ środowiska rodzinnego na relacje rodzinne

W celu zbadania, czy cechy środowiska rodzinnego wpływają na jakość relacji w rodzinie, przeprowadzono regresję logistyczną z binarną zmienną zależną - dobre relacje rodzinne (famrel) - oraz predyktorami takimi jak: wykształcenie matki (Medu), wykształcenie ojca (Fedu), status związku rodziców (Pstatus) oraz liczebność rodzinny (famsize).

Analiza wykazała, że jedynie status rodziców jest statystycznie istotnym predyktorem (Pstatus, $p = 0.0246$). Współczynnik dodatni dla tej zmiennej wskazuje, że uczniowie wychowujący się w pełnych rodzinach (gdzie rodzice żyją razem) mają większe prawdopodobieństwo deklarowania dobrych relacji rodzinnych. Może to

wynikać z większej stabilizacji emocjonalnej i obecności obojga opiekunów, co wpływa na poczucie bezpieczeństwa dziecka. W przypadku separacji lub rozwodu, dzieci często przeżywają konflikt lojalnościowy lub obwiniają jednego z rodziców, co może pogarszać relację z całą rodziną - a to znajduje odzwierciedlenie w odpowiedziach udzielanych w ankiecie.

Co ciekawe, poziom wykształcenia rodziców nie wykazał istotnego wpływu na jakość relacji rodzinnych, mimo że wyższe wykształcenie często wiąże się z bardziej wymagającą i angażującą karierą zawodową, co teoretycznie mogłoby ograniczać czas poświęcany dzieciom. Również liczliwość rodziny (famsize) nie okazała się istotna statystycznie - co sugeruje, że liczba członków rodziny nie musi wprost determinować bliskości relacji.

Dopasowanie modelu oceniono za pomocą statystyk deviance: wartość null deviance wynosiła 549.66, a residual deviance - 542.91, co oznacza nieznaczną poprawę dopasowania modelu względem modelu zerowego.

Podsumowując, relacje rodzinne uczniów okazały się niezależne od standardowych zmiennych społecznych, z wyjątkiem faktu wspólnego zamieszkania rodziców, co podkreśla znaczenie stabilności rodzinnej w percepcji domowego wsparcia emocjonalnego.

Test zgodności modelu (Hosmer-Lemeshow)

Aby ocenić dopasowanie modelu do danych empirycznych, zastosowano test Hosmera-Lemeshowa:

```
hoslem.test(model_famrel$y, fitted(model_famrel), g=10)
```

```
##  
##  Hosmer and Lemeshow goodness of fit (GOF) test  
##  
## data: model_famrel$y, fitted(model_famrel)  
## X-squared = 2.8959, df = 8, p-value = 0.9407
```

Ocena dopasowania modelu

Otrzymana wartość $p = 0.9407$ jest znacznie większa od poziomu istotności 0.05. Oznacza to, że nie ma podstaw do odrzucenia hipotezy o zgodności modelu z danymi. Model jest dobrze dopasowany – przewidywane prawdopodobieństwa nie odbiegają systematycznie od rzeczywistych wartości zmiennej zależnej.

Innymi słowy, model logistyczny uzyskuje akceptowalny poziom dopasowania i jego predykcje można uznać za wiarygodne w kontekście badania relacji rodzinnych na podstawie analizowanych predyktorów.

Etap 2: Wpływ relacji rodzinnych na zachowania społeczne ucznia

Kolejnym ogniwem analizowanego efektu domina są relacje rodzinne, które – oprócz bezpośredniego wpływu na dobrostan psychiczny ucznia – mogą również kształtować jego codzienne wybory i styl życia. W literaturze podkreśla się, że uczniowie doświadczający pozytywnych relacji w domu rzadziej przejawiają zachowania ryzykowne oraz częściej wykazują postawy prospołeczne.

W tej części analizie poddano zależność pomiędzy jakością relacji rodzinnych (operacyjonalizowaną za pomocą zmiennej famrel) a wybranymi zachowaniami społecznymi uczniów, takimi jak częstotliwość spotkań z przyjaciółmi (goout) oraz poziom spożycia alkoholu w weekendy i dni powszednie (Walc, Dalc). W celu uproszczenia interpretacji, relacje rodzinne sklasyfikowano binarnie: jako dobre (famrel ≥ 3) oraz negatywne jako pozostałe przypadki.

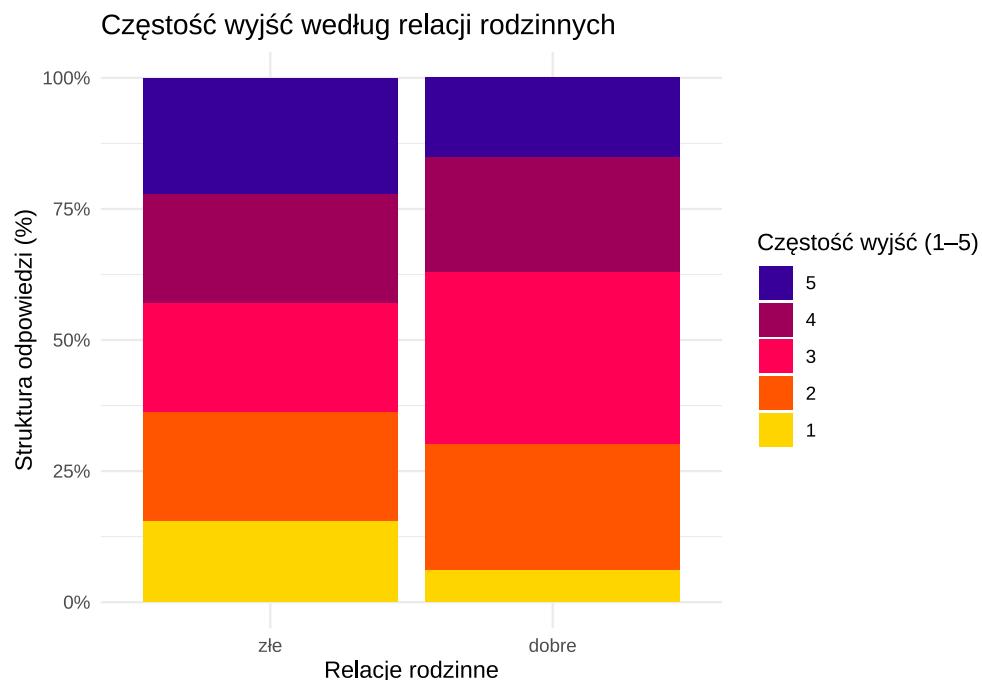
Celem tego etapu jest sprawdzenie, czy uczniowie deklarujący dobre relacje rodzinne rzadziej angażują się w intensywne życie towarzyskie oraz czy ograniczają spożywanie alkoholu. Odpowiedź na to pytanie pozwoli lepiej zrozumieć, czy relacje rodzinne mogą pełnić funkcję ochronną w procesie kształtowania stylu życia młodzieży oraz jakie mechanizmy stoją za tym zjawiskiem.

2.1. Relacje rodzinne a częstotliwość wyjść z przyjaciółmi

Ważnym aspektem codziennego funkcjonowania uczniów są kontakty towarzyskie, których częstotliwość może odzwierciedlać zarówno potrzeby społeczne, jak i wzorce wyniesione z domu. Istnieją przesłanki, że osoby wychowujące się w sprzyjającym środowisku rodzinnym rzadziej poszukują intensywnej integracji poza domem, ponieważ wsparcie i akceptację odnajdują w relacjach rodzinnych. Z kolei uczniowie z mniej satysfakcjonującymi więziami rodzinnymi mogą częściej szukać bliskości i zrozumienia wśród rówieśników.

W tej części analizie poddano związek pomiędzy jakością relacji rodzinnych (famrel_good) a częstotliwością wyjść z przyjaciółmi (goout). Celem jest ustalenie, czy pozytywne relacje w rodzinie ograniczają potrzebę częstego uczestnictwa w spotkaniach towarzyskich poza domem.

Wizualizacja rozkładu częstości wyjść



Interpretacja wykresu

Na wykresie widać, że uczniowie z dobrymi relacjami rodzinnymi rzadziej deklarują bardzo częste wyjścia z przyjaciółmi (poziom 5) niż ich rówieśnicy z mniej satysfakcjonującym zapleczem domowym. Sugeruje to, że silne więzi rodzinne mogą zmniejszać potrzebę intensywnej integracji towarzyskiej – uczniowie ci znajdują wsparcie i bliskość w domu, przez co rzadziej szukają ich poza nim. Co ciekawe, wśród osób z gorszymi relacjami rodzinnymi częściej obserwuje się zarówno bardzo wysoką, jak i bardzo niską częstotliwość wyjść. Może to świadczyć o dwóch skrajnych reakcjach na deficyty w środowisku domowym – jedni kompensują je nadmierną aktywnością społeczną, inni wycofują się i izolują.

Modelowanie statystyczne

Aby zweryfikować powyższą zależność, zastosowano model regresji Poissona. Model uwzględniał także dodatkowe zmienne kontrolne, takie jak płeć, wiek, ilość wolnego czasu oraz miejsce zamieszkania.

```
model_goout <- glm(goout ~ famrel_good + sex + age + freetime + address,
                     family = "poisson", data = students)
summary(model_goout)

##
## Call:
## glm(formula = goout ~ famrel_good + sex + age + freetime + address,
##       family = "poisson", data = students)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.167538  0.251840  0.665   0.506
## famrel_good -0.009367  0.066996 -0.140   0.889
## sexM        0.004622  0.035738  0.129   0.897
## age         0.034872  0.013886  2.511   0.012 *
## freetime    0.115667  0.017446  6.630 3.36e-11 ***
## addressU    0.036885  0.039456  0.935   0.350
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 462.51  on 1043  degrees of freedom
## Residual deviance: 409.34  on 1038  degrees of freedom
## AIC: 3522.5
##
## Number of Fisher Scoring iterations: 4
```

Wyniki modelu: wpływ relacji rodzinnych na częstotliwość wyjść z przyjaciółmi

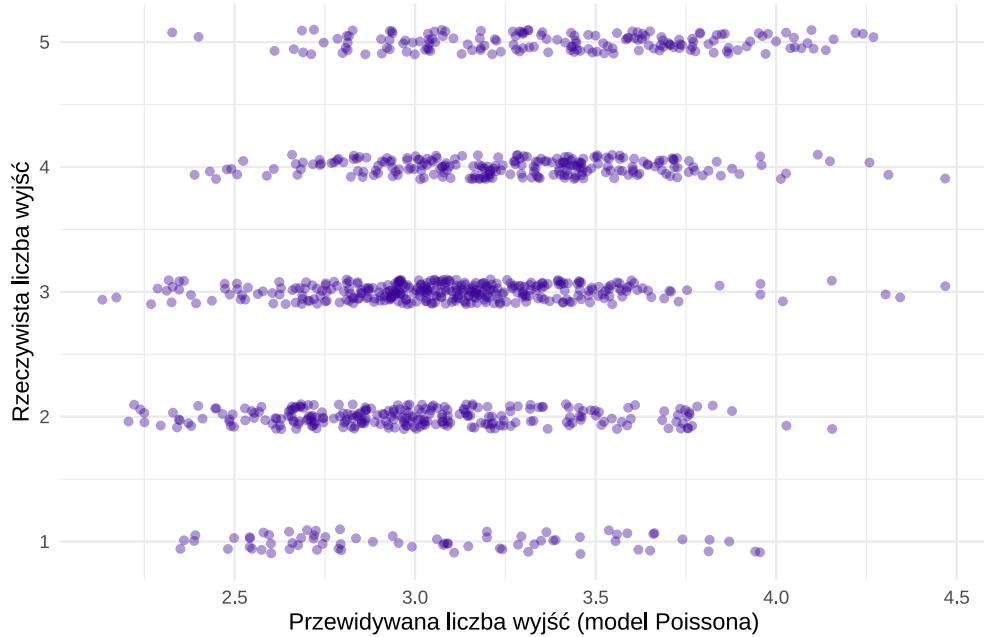
Analiza wykazała, że jedynymi istotnymi predyktorami częstotliwości wyjść są wiek ucznia (age, $p = 0.012$) oraz ilość wolnego czasu (freetime, $p < 0.001$). Oznacza to, że starsi uczniowie oraz osoby mające więcej czasu wolnego częściej wychodzą z przyjaciółmi, co jest intuicyjne i zgodne z oczekiwaniemi. Wpływ płci i miejsca zamieszkania nie okazał się statystycznie istotny.

Co istotne, dobra jakość relacji rodzinnych nie wpływa istotnie statystycznie na częstotliwość wyjść (famrel_good, $p = 0.889$). Otrzymany współczynnik jest praktycznie nieistotny, co sugeruje, że niezależnie od tego, czy uczeń deklaruje dobre relacje rodzinne, częstotliwość kontaktów towarzyskich pozostaje podobna.

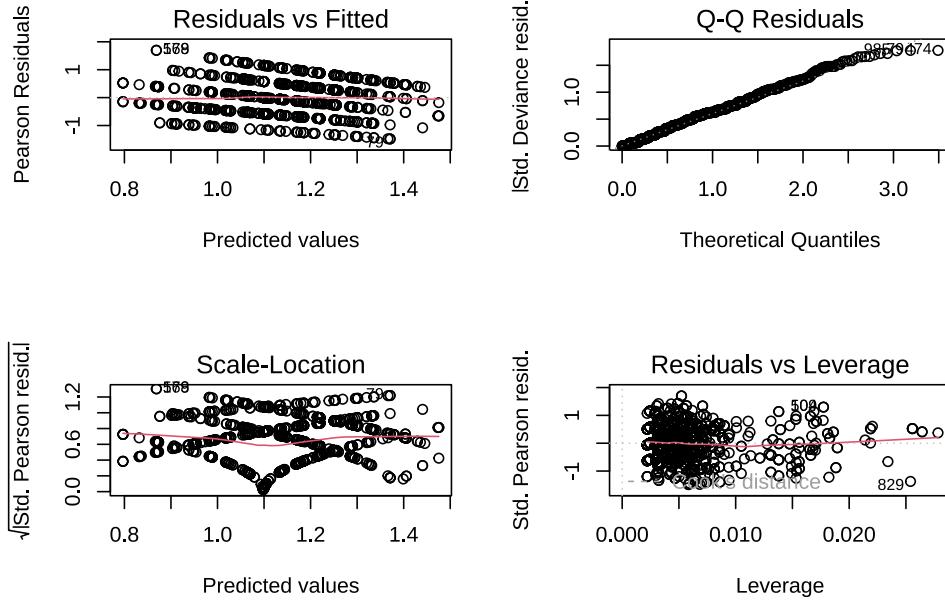
Dopasowanie modelu oceniono za pomocą statystyk deviance: wartość null deviance wynosiła 462.51, a residual deviance - 409.34, co wskazuje na zauważalną poprawę względem modelu zerowego. Model dobrze oddaje wpływ wieku i czasu wolnego, ale nie potwierdza istotnej roli relacji rodzinnych w zakresie intensywności życia towarzyskiego.

Podsumowując, wyniki analizy sugerują, że kluczowymi czynnikami sprzyjającymi częstym wyjściom z przyjaciółmi są wiek oraz ilość wolnego czasu, natomiast dobra relacja z rodziną nie ogranicza ani nie wzmacnia skłonności do aktywności społecznej w tym wymiarze. Wynik ten może świadczyć o tym, że potrzeby kontaktu z rówieśnikami realizowane są niezależnie od sytuacji rodzinnej, a oparte są głównie na zasobach czasowych i dojrzałości ucznia.

Rzeczywiste vs przewidywane wyjścia z przyjaciółmi



Wyniki modelu i diagnostyka



Ocena dopasowania modelu

W celu oceny jakości dopasowania modelu Poissona, opisującego zależność pomiędzy jakością relacji rodzinnych a częstotliwością wyjść z przyjaciółmi, przeprowadzono analizę diagnostyczną opartą na standardowych wykresach reszt. Analiza ta pozwala ocenić, czy model prawidłowo odwzorowuje strukturę danych i spełnia założenia niezbędne do rzetelnej interpretacji wyników.

Na wykresie *Residuals vs Fitted* reszty rozmiieszczone są w sposób losowy wokół zera, bez wyraźnych wzorców czy systematycznych odchyлеń. Sugeruje to, że model nie popełnia systematycznych błędów dla określonych wartości przewidywanych i nie występuje poważny problem z nielinowością czy niedopasowaniem.

Wykres *Q-Q Residuals* ukazuje pewne odchylenia od idealnego rozkładu normalnego, szczególnie w wyższych kwantylach, jednak dla modeli Poissona umiarkowane odchylenia są typowe i nie dyskwalifikują modelu – istotne byłoby tu pojawienie się silnych „ogonów” lub bardzo dużych odchyleń, których na wykresie nie obserwujemy w istotnym stopniu.

Na wykresie *Scale-Location* rozrzut reszt względem wartości przewidywanych nie wykazuje silnego trendu wzrostowego, co oznacza, że wariancja reszt pozostaje względnie stała dla różnych poziomów zmiennej zależnej. Brak wyraźnej heteroskedastyczności wzmacnia wiarygodność modelu.

Wreszcie, wykres *Residuals vs Leverage* nie wskazuje na obecność pojedynczych obserwacji o wyjątkowo dużej dźwigni czy nadmiernym wpływie na dopasowanie modelu. Większość punktów mieści się w bezpiecznym zakresie, nie pojawiają się skrajnie wpływowe obserwacje (Cook's distance).

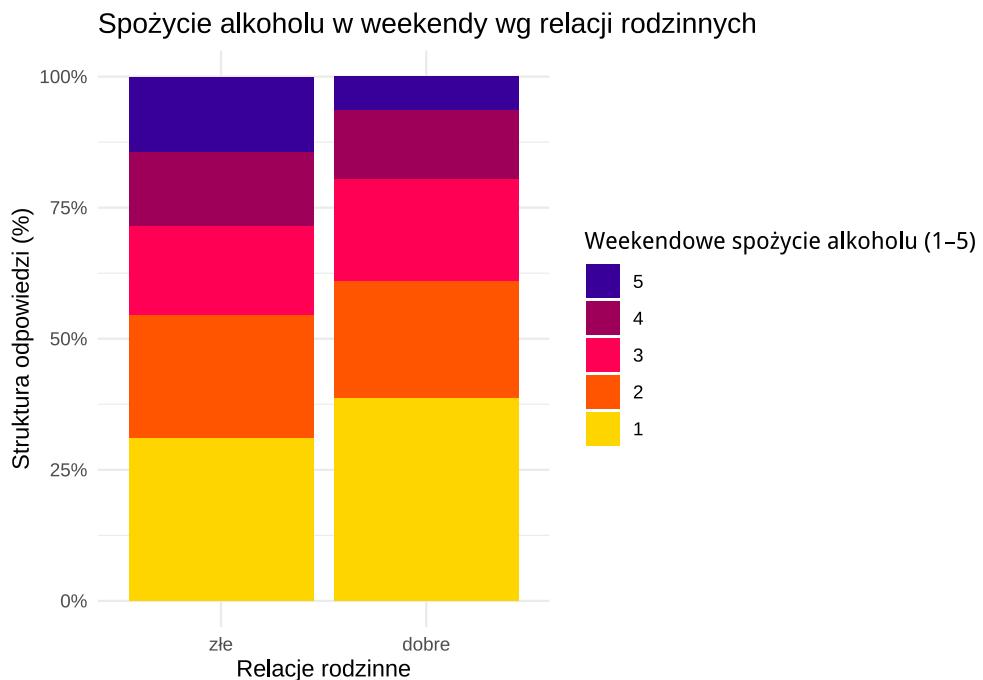
Podsumowując, model Poissona został poprawnie dopasowany do danych. Analiza diagnostyczna nie wykazuje poważnych naruszeń założeń ani anomalii, które uniemożliwiałyby rzetelną interpretację uzyskanych wyników. Otrzymane rezultaty można zatem uznać za wiarygodne w kontekście przyjętej specyfikacji modelu.

2.2. Relacje rodzinne a spożycie alkoholu w weekendy

Kolejnym elementem zachowań społecznych, potencjalnie zależnym od relacji rodzinnych, jest spożycie alkoholu. Częstsze sięganie po alkohol może być formą radzenia sobie ze stresem lub próbą integracji z grupą rówieśniczą.

W tej części analizie poddano związek między jakością relacji rodzinnych a poziomem spożycia alkoholu w weekendy (Walc).

Wizualizacja rozkładu spożycia alkoholu w weekendy



Interpretacja wykresu

Widzimy, że wśród uczniów z dobrymi relacjami rodzinymi odsetek osób deklarujących bardzo wysokie spożycie alkoholu w weekendy (poziom 5) jest niższy niż wśród uczniów z gorszymi relacjami. Potwierdza to tezę, że dobre relacje rodzinne mogą być czynnikiem chroniącym przed nadmiernym spożyciem alkoholu.

Modelowanie statystyczne

W analizie zastosowano model regresji Poisson'a:

```
model_walc <- glm(Walc ~ famrel_good + sex + age + goout + freetime,
                     family = "poisson", data = students)
summary(model_walc)
```

```
## 
## Call:
## glm(formula = Walc ~ famrel_good + sex + age + goout + freetime,
##      family = "poisson", data = students)
## 
```

```

## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.20454   0.29087 -0.703   0.4819
## famrel_good -0.13212   0.07431 -1.778   0.0754 .
## sexM         0.32345   0.04169  7.759 8.56e-15 ***
## age          0.02759   0.01649  1.673   0.0944 .
## goout        0.18943   0.01912  9.907 < 2e-16 ***
## freetime     -0.02661   0.02132 -1.248   0.2119
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 734.57 on 1043 degrees of freedom
## Residual deviance: 549.22 on 1038 degrees of freedom
## AIC: 3274.6
##
## Number of Fisher Scoring iterations: 4

```

Wyniki modelu: wpływ relacji rodzinnych i czynników demograficznych na spożycie alkoholu w weekendy

W celu zbadania, jakie czynniki wpływają na częstotliwość spożycia alkoholu w weekendy, przeprowadzono regresję Poissona, w której zmienną zależną była częstość spożycia alkoholu w weekendy (Walc). Do modelu włączono jakość relacji rodzinnych (famrel_good), płeć (sex), wiek (age), częstotliwość wyjść z przyjaciółmi (goout) oraz ilość wolnego czasu (freetime).

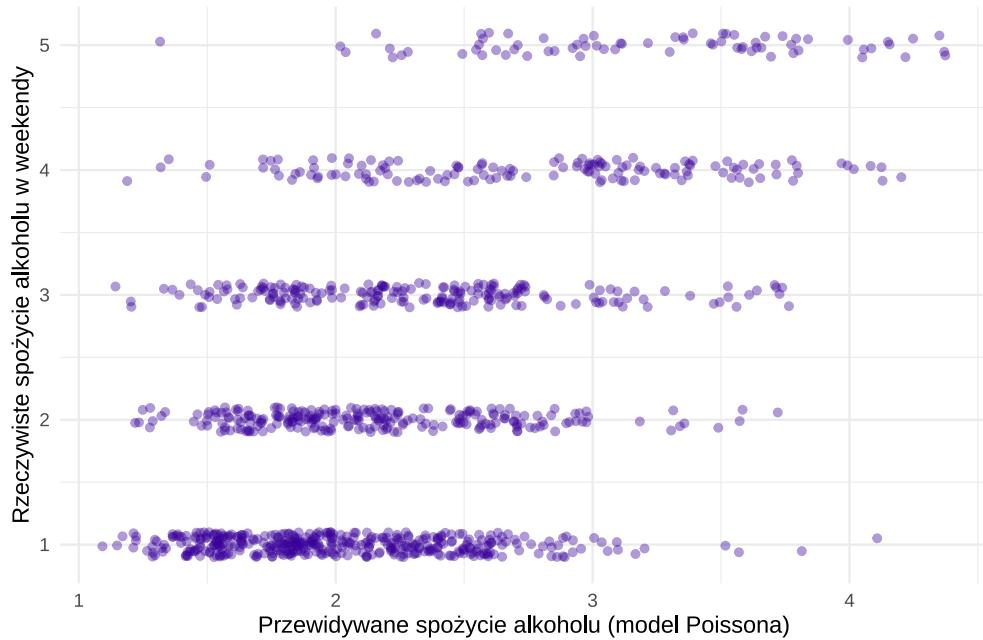
Analiza wykazała, że kluczowymi czynnikami sprzyjającymi częstszemu spożyciu alkoholu w weekendy są płeć oraz intensywność życia towarzyskiego. Chłopcy istotnie częściej sięgają po alkohol niż dziewczęta (sexM, $p < 0.001$), a wyższa częstotliwość wyjść z przyjaciółmi (goout, $p < 0.001$) także silnie zwiększa prawdopodobieństwo spożywania alkoholu. Pozostałe zmienne (wiek oraz ilość wolnego czasu) nie osiągnęły poziomu istotności statystycznej.

Jeśli chodzi o relacje rodzinne, ujemny współczynnik dla zmiennej famrel_good sugeruje, że dobre relacje rodzinne mogą wiązać się z nieco rzadszym spożywaniem alkoholu, jednak ten efekt jest na granicy istotności statystycznej (famrel_good, $p = 0.0754$). Oznacza to, że wpływ wsparcia rodzinnego jest zauważalny, ale w analizowanej grupie nie osiąga tradycyjnego poziomu istotności ($p < 0.05$).

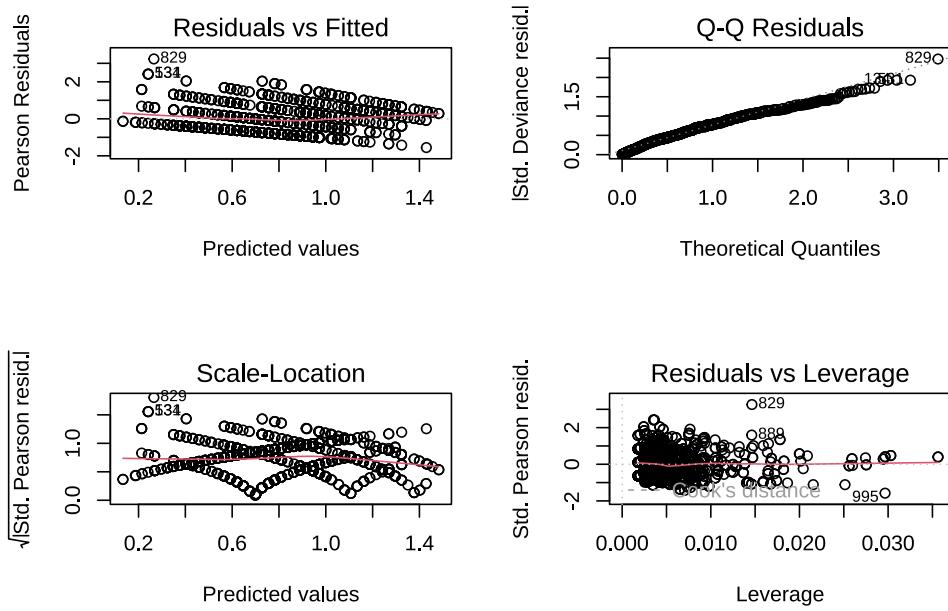
Dopasowanie modelu oceniono za pomocą statystyk deviance: null deviance wynosiła 734.57, residual deviance - 549.22, co oznacza istotną poprawę względem modelu zerowego.

Podsumowując, wyniki potwierdzają, że spożywanie alkoholu w weekendy jest przede wszystkim związane z płcią oraz aktywnością towarzyską uczniów, podczas gdy wpływ jakości relacji rodzinnych – choć negatywny – pozostaje na pograniczu istotności. Sugeruje to, że interwencje ograniczające ryzykowne zachowania powinny koncentrować się zwłaszcza na grupach o wysokiej aktywności towarzyskiej oraz chłopcach.

Rzeczywiste vs przewidywane spożycie alkoholu w weekendy



Wyniki modelu i diagnostyka



Ocena dopasowania modelu

W celu oceny jakości dopasowania modelu Poissona, opisującego zależność pomiędzy analizowanymi czynnikami a częstotliwością spożycia alkoholu w weekendy (Walc), przeprowadzono analizę diagnostyczną opartą na standardowych wykresach reszt. Pozwala to zweryfikować, czy model prawidłowo odwzorowuje strukturę danych i czy spełnione są kluczowe założenia statystyczne.

Na wykresie *Residuals vs Fitted* większość reszt skupia się wokół zera, jednak widoczne są obserwacje o wyższych wartościach reszt, które mogą sugerować pewne niedopasowanie modelu w niektórych przedziałach przewidywanych wartości. Większość danych nie wykazuje jednak wyraźnych wzorców czy systematycznych błędów predykcji.

Wykres *Q-Q Residuals* wskazuje na umiarkowane odchylenia rozkładu reszt od rozkładu teoretycznego, zwłaszcza w górnym kwantylach. Jest to jednak typowe dla danych o ograniczonej skali i nie stanowi poważnego naruszenia dla interpretacji globalnych trendów.

Wykres *Scale-Location* nie wykazuje wyraźnego wzrostu zmienności reszt wraz z wartościami przewidywanymi - sugeruje to względną stabilność wariancji w całym zakresie modelowanych wartości, a więc nie występuje istotny problem heteroskedastyczności.

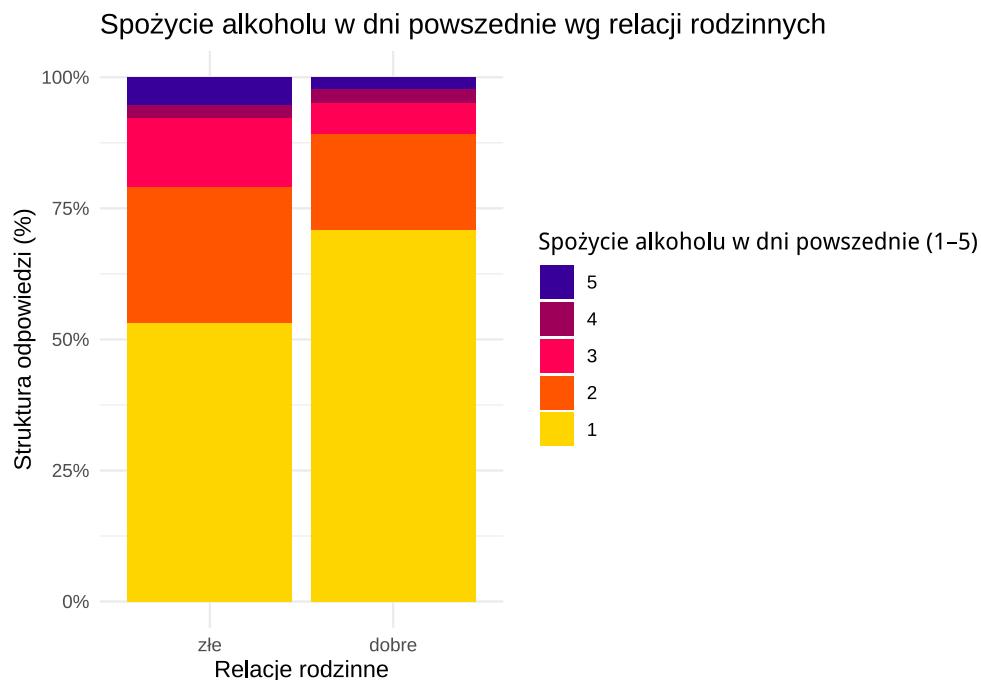
Na wykresie *Residuals vs Leverage* można zauważać kilka obserwacji o nieco większej dźwigni, lecz żadna nie przekracza umownych progów skrajnego wpływu. Nie ma wyraźnych obserwacji dominujących (Cook's distance).

Podsumowując, analiza diagnostyczna nie wskazuje na istotne naruszenia założeń modelu Poissona. Model można uznać za poprawnie dopasowany i jego wyniki są wiarygodne w kontekście analizy wpływu czynników rodzinnych i demograficznych na częstość spożywania alkoholu w weekendy przez uczniów.

2.3. Relacje rodzinne a spożycie alkoholu w dni powszednie

Analogicznie przeanalizowano zależność pomiędzy jakością relacji rodzinnych a spożyciem alkoholu w dni powszednie (Dalc).

Wizualizacja rozkładu spożycia alkoholu w dni powszednie



Interpretacja wykresu

Widzimy, że wśród uczniów deklarujących dobre relacje rodzinne odsetek osób zgłaszających bardzo wysokie spożycie alkoholu w dni powszednie (poziom 5) jest wyraźnie niższy niż w grupie uczniów z gorszymi relacjami. Zdecydowana większość uczniów z pozytywnymi więzami rodzinnymi wskazuje na niskie lub umiarkowane spożycie alkoholu, podczas gdy wśród osób z mniejszą satysfakcjonującymi relacjami częściej pojawiają się odpowiedzi o wyższym spożyciu. Potwierdza to, że dobre relacje rodzinne mogą pełnić funkcję czynnika ochronnego także w kontekście codziennego, a nie tylko weekendowego, picia alkoholu.

Modelowanie statystyczne

W analizie zastosowano model regresji Poisson'a:

```
model_dalc <- glm(Dalc ~ famrel_good + sex + age + goout + freetime,
                     family = "poisson", data = students)
summary(model_dalc)
```

```
##  
## Call:  
## glm(formula = Dalc ~ famrel_good + sex + age + goout + freetime,  
##       family = "poisson", data = students)  
##
```

```

## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.90589   0.35505 -2.551   0.0107 *
## famrel_good -0.22213   0.08900 -2.496   0.0126 *
## sexM         0.31956   0.05159  6.194 5.87e-10 ***
## age          0.05612   0.02013  2.788   0.0053 **
## goout        0.11076   0.02338  4.737 2.17e-06 ***
## freetime     0.01814   0.02642  0.686   0.4924
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 451.98 on 1043 degrees of freedom
## Residual deviance: 362.22 on 1038 degrees of freedom
## AIC: 2722.4
##
## Number of Fisher Scoring iterations: 4

```

Wyniki modelu: wpływ relacji rodzinnych i czynników demograficznych na spożycie alkoholu w dni powszednie

W celu zbadania, jakie czynniki wpływają naczęstość spożycia alkoholu w dni powszednie, przeprowadzono regresję Poissona ze zmienną zależną opisującą spożycie alkoholu w dni powszednie (Dalc). Do modelu włączono: jakość relacji rodzinnych (famrel_good), płeć (sex), wiek (age), częstotliwość wyjścia z przyjaciółmi (goout) oraz ilość wolnego czasu (freetime).

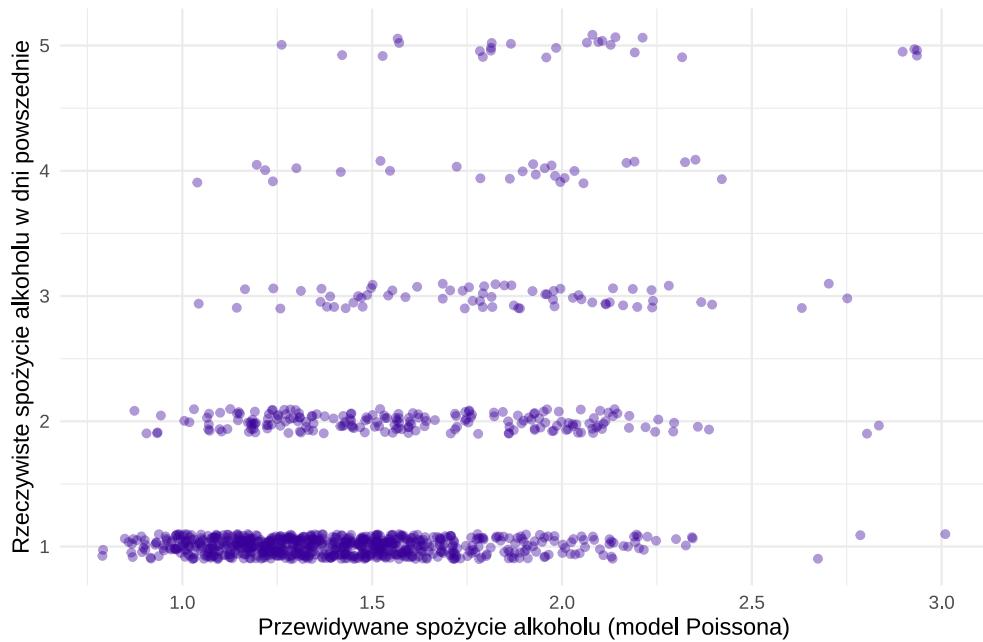
Analiza wykazała, że wszystkie kluczowe zmienne z wyjątkiem wolnego czasu mają istotny wpływ naczęstość spożywania alkoholu w dni powszednie. Po pierwsze, dobre relacje rodzinne istotnie obniżają ryzyko częstego spożywania alkoholu (famrel_good, $p = 0.0126$) – uczniowie deklarujący pozytywne relacje w rodzinie rzadziej sięgają po alkohol w ciągu tygodnia. Wpływ ten jest silniejszy i bardziej jednoznaczny niż w przypadku spożycia alkoholu w weekendy.

Po drugie, płeć (chłopcy częściej niż dziewczęta, $p < 0.001$), wyższy wiek ($p = 0.0053$) oraz częstsze wyjścia z przyjaciółmi (goout, $p < 0.001$) również istotnie zwiększają prawdopodobieństwo sięgania po alkohol w dni robocze. Ilość wolnego czasu nie okazała się predyktorem istotnym statystycznie.

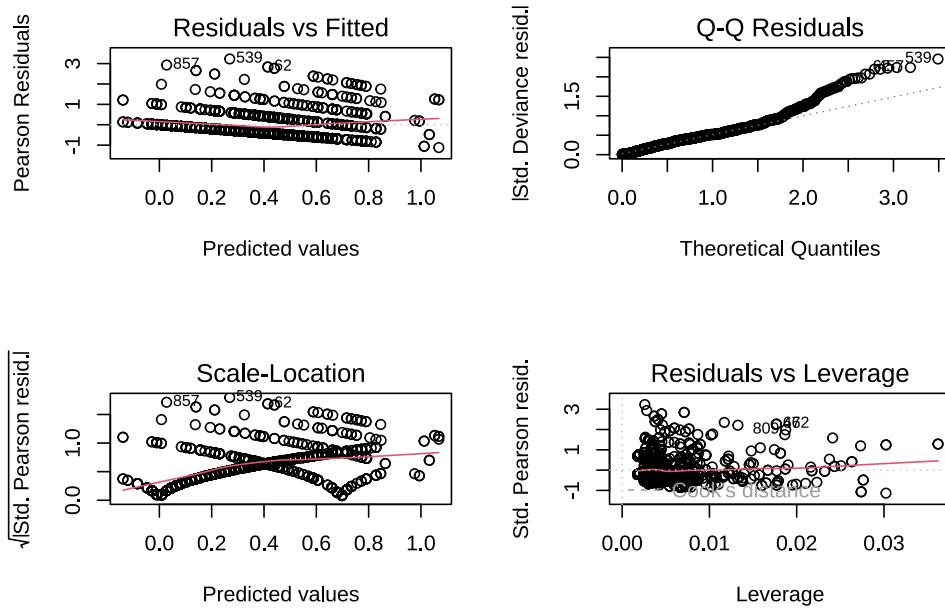
Dopasowanie modelu oceniono na podstawie statystyk deviance: null deviance wynosiła 451.98, a residual deviance – 362.22, co świadczy o wyraźnej poprawie względem modelu zerowego.

Podsumowując, dobre relacje rodzinne odgrywają wyraźną rolę ochronną przed sięganiem po alkohol w dni powszednie, co pokazuje, że wsparcie ze strony rodziny może być szczególnie ważne w kontekście codziennych, a nie tylko okazjonalnych, zachowań ryzykownych. Dodatkowo, wyższe ryzyko dotyczy chłopców, starszych uczniów oraz osób bardziej aktywnych towarzysko. Wyniki te sugerują, że działania profilaktyczne i wychowawcze warto kierować szczególnie do tych grup, a wzmacnianie więzi rodzinnych może być skuteczną strategią ograniczania ryzykownych zachowań młodzieży.

Rzeczywiste vs przewidywane spożycie alkoholu w dni powszednie



Wyniki modelu i diagnostyka



Ocena dopasowania modelu

W celu oceny jakości dopasowania modelu Poissona, opisującego zależność pomiędzy analizowanymi czynnikami a częstością spożycia alkoholu w dni powszednie (Dalc), przeprowadzono analizę diagnostyczną opartą na standardowych wykresach reszt. Pozwala to zweryfikować, czy model prawidłowo odwzorowuje strukturę danych i czy spełnione są kluczowe założenia statystyczne.

Na wykresie \textit{Residuals vs Fitted} większość reszt skupia się wokół zera, jednak widoczne są obserwacje o wyższych wartościach reszt, co może sugerować pewne lokalne niedopasowanie modelu w wybranych przedziałach wartości przewidywanych. Nie widać jednak wyraźnych, systematycznych odchyleń, które wskazywałyby na poważny problem z predykcją.

Wykres \textit{Q-Q Residuals} ukazuje umiarkowane odchylenia rozkładu reszt od teoretycznego rozkładu normalnego, szczególnie w wyższych kwantylach. Jest to typowe dla danych o ograniczonej skali i nie stanowi poważnego naruszenia dla interpretacji globalnych trendów.

Wykres \textit{Scale-Location} nie wykazuje wyraźnego wzrostu zmienności reszt wraz ze wzrostem wartości przewidywanych, co sugeruje względną stabilność wariancji w całym zakresie modelowanych wartości i brak istotnych problemów z heteroskedastycznością.

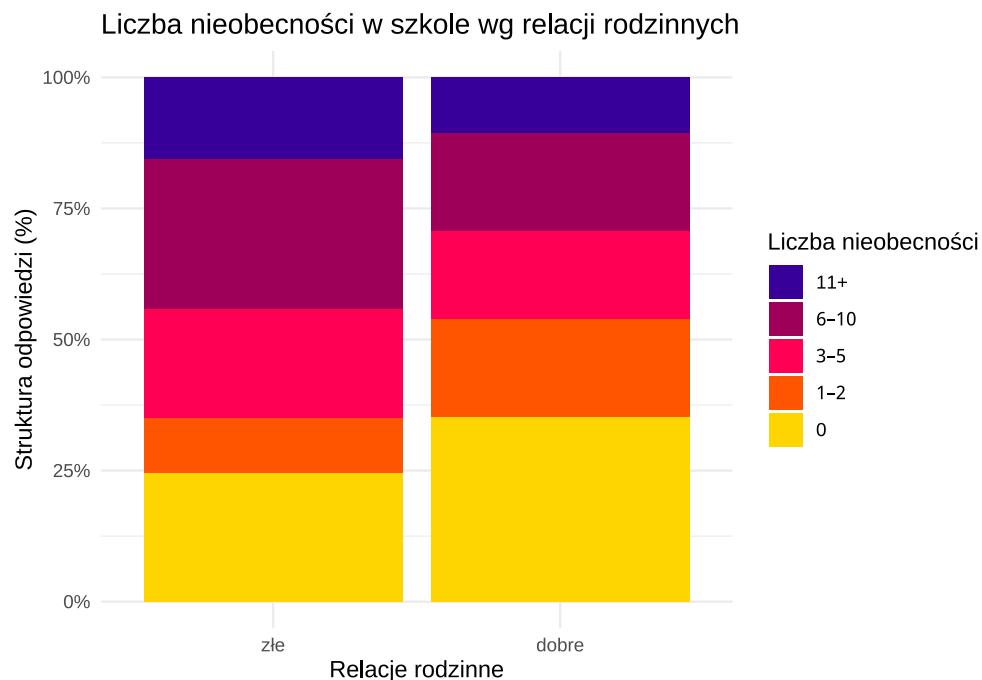
Na wykresie \textit{Residuals vs Leverage} można zaobserwować kilka punktów o nieco wyższej dźwigni, ale żaden z nich nie przekracza przyjętych progów skrajnego wpływu (Cook's distance). Nie występują wyraźnie dominujące obserwacje.

Podsumowując, analiza diagnostyczna nie wskazuje na istotne naruszenia założeń modelu Poissona. Model można uznać za poprawnie dopasowany i jego wyniki są wiarygodne w kontekście analizy wpływu czynników rodzinnych oraz demograficznych naczęstość spożywania alkoholu w dni powszednie przez uczniów.

2.4. Relacje rodzinne a nieobecności

Analogicznie przeanalizowano zależność pomiędzy jakością relacji rodzinnych a nieobecnościami.

Wizualizacja rozkładu nieobecności



Interpretacja wykresu

Na wykresie widać wyraźną różnicę w strukturze nieobecności pomiędzy uczniami deklarującymi dobre i złe relacje rodzinne. W grupie uczniów z pozytywnymi więzami rodzinnymi częściej występują deklaracje zerowej lub bardzo niskiej liczby nieobecności, natomiast wśród osób z gorszymi relacjami wyraźnie częstsze są odpowiedzi wskazujące na większą liczbę opuszczonych dni. Może to sugerować, że stabilne i wspierające środowisko domowe sprzyja systematycznemu uczestnictwu w zajęciach szkolnych, pełniąc funkcję ochronną przed absencją, która może się wiązać z niższymi wynikami edukacyjnymi.

Modelowanie statystyczne

W analizie zastosowano model regresji Poisson'a:

```
model_abs <- glm(absences ~ famrel_good + sex + age + goout + freetime + Walc + Dalc,  
                    family = "poisson", data = students)  
summary(model_abs)
```

```
##  
## Call:  
## glm(formula = absences ~ famrel_good + sex + age + goout + freetime +  
##       Walc + Dalc, family = "poisson", data = students)  
##  
## Coefficients:
```

```

##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.9804241  0.2041480 -4.803 1.57e-06 ***
## famrel_good -0.1361406  0.0517764 -2.629  0.00855 **
## sexM        -0.1613909  0.0328411 -4.914 8.91e-07 ***
## age          0.1456832  0.0116111 12.547 < 2e-16 ***
## goout        -0.0005003  0.0147929 -0.034  0.97302
## freetime     -0.0635748  0.0152119 -4.179 2.92e-05 ***
## Walc         0.1230030  0.0153895  7.993 1.32e-15 ***
## Dalc         0.0733533  0.0183154  4.005 6.20e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 6914.0 on 1043 degrees of freedom
## Residual deviance: 6499.3 on 1036 degrees of freedom
## AIC: 8897.8
##
## Number of Fisher Scoring iterations: 6

```

Wyniki modelu: wpływ relacji rodzinnych i czynników demograficznych na spożycie alkoholu w dni powszednie

W celu identyfikacji czynników wpływających naczęstość nieobecności uczniów w szkole przeprowadzono analizę regresji Poissona z liczbą nieobecności (absences) jako zmienną zależną. W modelu uwzględniono: jakość relacji rodzinnych (famrel_good), płeć (sex), wiek (age), częstotliwość wyjścia z przyjaciółmi (goout), ilość wolnego czasu (freetime) oraz spożycie alkoholu zarówno w tygodniu (Dalc), jak i w weekendy (Walc).

Analiza wykazała, że niemal wszystkie predyktory – z wyjątkiem goout – mają istotny statystycznie wpływ na liczbę nieobecności. Po pierwsze, dobre relacje rodzinne wyraźnie ograniczają liczbę opuszczonych zajęć (famrel_good, $p = 0.0085$), co potwierdza ich rolę jako czynnika ochronnego, wspierającego regularne uczestnictwo w nauce.

Ciekawym wynikiem jest wpływ płci – chłopcy, choć częściej sięgają po alkohol, mają przeciętnie mniej nieobecności niż dziewczęta (sexM, $p < 0.001$). Może to wynikać z odmiennych strategii radzenia sobie ze stresem lub z różnic w podejściu do obowiązków szkolnych.

Z kolei wiek ucznia (age, $p < 0.001$) istotnie zwiększa liczbę nieobecności – starsi uczniowie częściej pozwalają sobie na opuszczanie zajęć, co może być oznaką rosnącej autonomii, zmęczenia edukacją lub większej presji z innych obszarów życia (np. pracy).

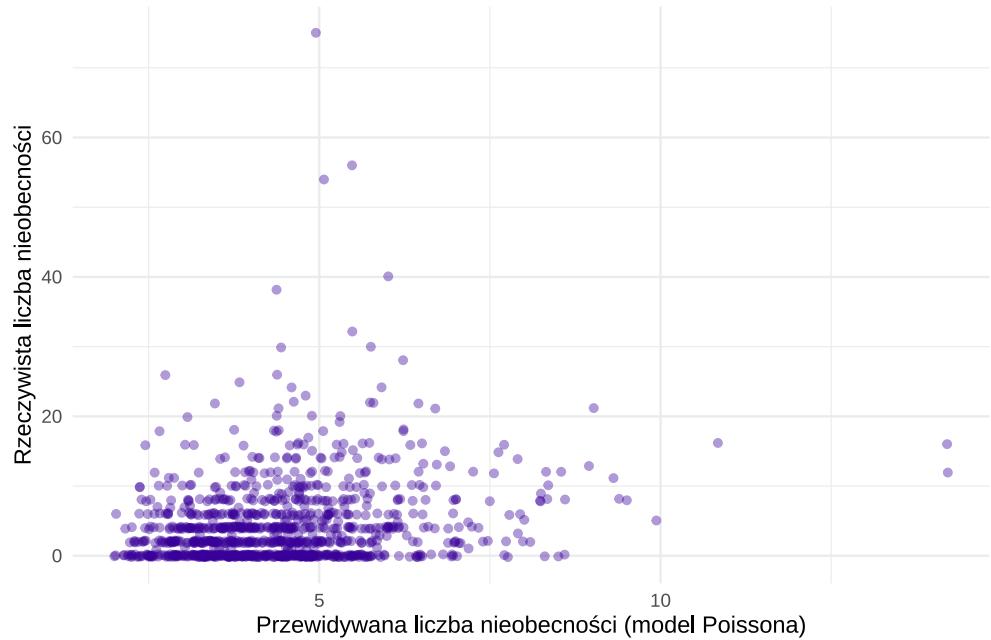
Ilość wolnego czasu (freetime, $p < 0.001$) ma odwrotny, ochronny wpływ – im więcej wolnych godzin, tym rzadziej uczniowie decydują się na absencję. Można to interpretować jako efekt kompensacyjny: uczniowie, którzy mają mało czasu wolnego, mogą uciekać się do nieobecności w szkole, by ten czas odzyskać.

Oba wskaźniki spożycia alkoholu – zarówno (Walc, $p < 0.001$), jak i (Dalc, $p < 0.001$) – silnie korelują z liczbą nieobecności. To potwierdza, że absencje mogą stanowić przejaw zachowań buntowniczych, a osoby sięgające po alkohol są jednocześnie bardziej skłonne do nieregularnego uczęszczania na zajęcia.

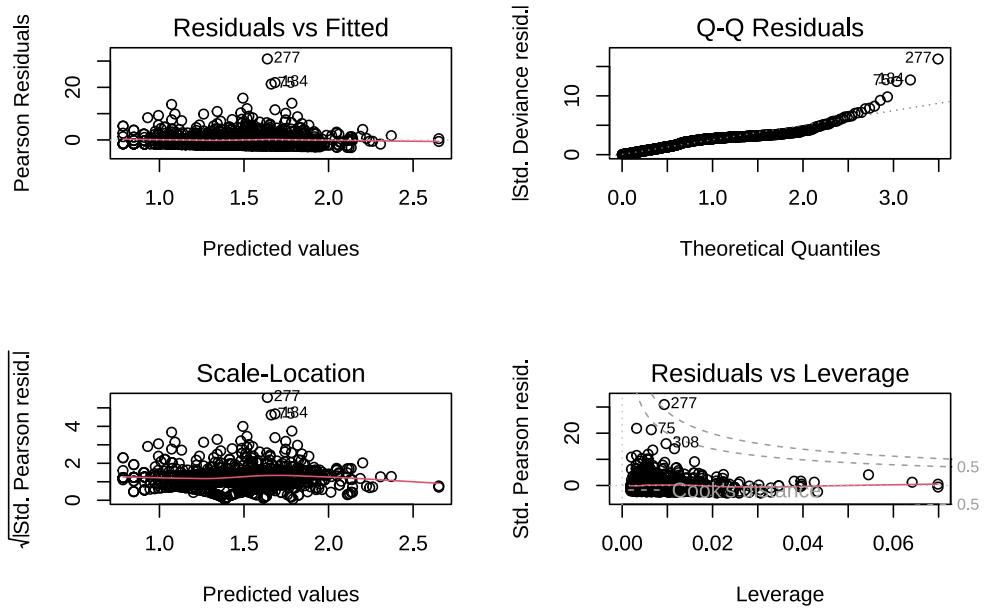
Jedyną zmienną, która nie osiągnęła poziomu istotności statystycznej, jest (goout, $p = 0.97$) – częstotliwość wyjścia z przyjaciółmi nie wpływa na liczbę nieobecności, co może wskazywać, że to nie sama aktywność towarzyska, lecz sposób jej organizacji i pozostałe aspekty stylu życia mają większe znaczenie.

Podsumowując, model potwierdza, że nieobecności w szkole są silnie powiązane ze stylem życia uczniów, a także z ich relacjami rodzinnymi i postawą wobec obowiązków. Wyniki te podkreślają konieczność wspierania uczniów w budowaniu zdrowych nawyków, a także uwzględniania kontekstu rodzinnego i emocjonalnego w działaniach wychowawczych i profilaktycznych.

Rzeczywiste vs przewidywane nieobecności uczniów



Wyniki modelu i diagnostyka



Ocena dopasowania modelu

Aby ocenić jakość dopasowania modelu Poissona prognozującego liczbę nieobecności uczniów, przeprowadzono analizę diagnostyczną opartą na standardowych wykresach reszt.

Na wykresie Residuals vs Fitted wyraźnie widać, że model systematycznie niedoszacowuje liczbę nieobecności w przypadku wielu uczniów – dla niektórych przewidywania odbiegają nawet o ponad 30 jednostek od wartości rzeczywistych. Świadczy to o poważnym lokalnym niedopasowaniu i wskazuje na ograniczoną zdolność modelu do uchwycenia ekstremalnych przypadków absencji.

Wykres Q-Q Residuals ukazuje znaczne odchylenia reszt od rozkładu normalnego – szczególnie w środkowej części rozkładu. Taki rozkład reszt sugeruje, że założenie normalności jest wyraźnie naruszone, co może wpływać na wiarygodność wartości p i interpretację efektów w modelu.

Na wykresie Scale-Location nie obserwuje się istotnego wzrostu wariancji reszt w zależności od wartości przewidywanych – układ punktów jest względnie równomierny, co nie wskazuje na poważny problem z heteroskedastycznością.

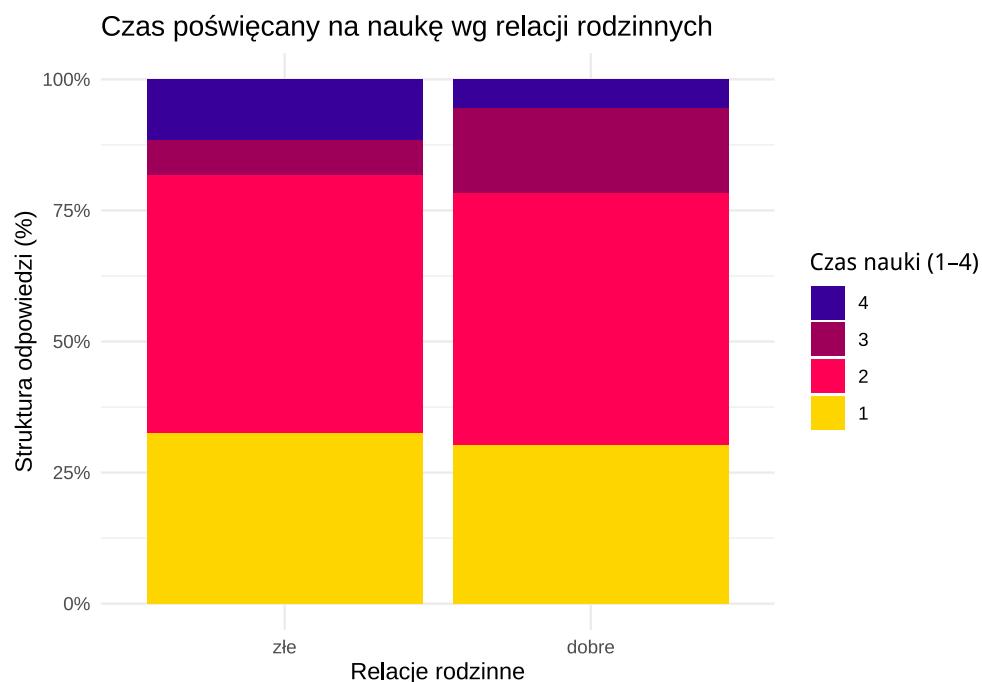
Wreszcie, wykres Residuals vs Leverage pokazuje kilka potencjalnie wpływowych obserwacji, z których jedna – przypadek nr 277 – przekracza wartość 1 dla dystansu Cooka. Może to oznaczać, że pojedyncze przypadki o bardzo wysokiej liczbie nieobecności znacząco wpływają na dopasowanie modelu.

Podsumowując: mimo że model identyfikuje istotne zależności między predyktorami a absencją, jego dopasowanie do skrajnych przypadków pozostawia wiele do życzenia. Niedoszacowanie w górnym zakresie oraz naruszenie rozkładu reszt sugerują, że model Poissona może nie być wystarczający do pełnego uchwycenia mechanizmu absencji w populacji uczniów.

2.5. Relacje rodzinne a ilość nauki

Na koniec tego etapu zależność pomiędzy jakością relacji rodzinnych a czasem poświęconym na naukę.

Wizualizacja rozkładu nieobecności



Interpretacja wykresu

Na wykresie przedstawiono czas poświęcany na naukę w zależności od jakości relacji rodzinnych. Choć ogólna struktura odpowiedzi jest zbliżona, można zauważać, że wśród uczniów z gorszymi relacjami rodzinnymi nieco częściej pojawiają się deklaracje intensywnej nauki (najwyższa kategoria). Może to sugerować, że część tych osób kompensuje brak wsparcia emocjonalnego większym zaangażowaniem w naukę. Tego typu zachowanie – polegające na zamknięciu się w sobie i skupieniu na edukacji – może być formą ucieczki od trudności rodzinnych. Mimo to, trudno mówić o jednoznacznej zależności – różnice nie są wyraźne, a większość uczniów w obu grupach deklaruje umiarkowany czas nauki.

Modelowanie statystyczne

W analizie zastosowano model regresji Poisson'a:

```
model_studytime <- glm(studytme ~ famrel_good + absences + freetime + sex + Medu + Fedu,
                         family = "poisson", data = students)
summary(model_studytime)
```

```
##  
## Call:  
## glm(formula = studytme ~ famrel_good + absences + freetime +  
##       sex + Medu + Fedu, family = "poisson", data = students)  
##
```

```

## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.739270  0.119713  6.175 6.60e-10 ***
## famrel_good 0.012574  0.084510  0.149  0.8817
## absences    -0.006123  0.003740 -1.637  0.1016
## freetime     -0.021666  0.021827 -0.993  0.3209
## sexM        -0.212596  0.046372 -4.585 4.55e-06 ***
## Medu         0.055857  0.025720  2.172  0.0299 *
## Fedu        -0.015458  0.026093 -0.592  0.5536
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 360.00 on 1043 degrees of freedom
## Residual deviance: 329.32 on 1037 degrees of freedom
## AIC: 2979.2
##
## Number of Fisher Scoring iterations: 4

```

Wyniki modelu: wpływ relacji rodzinnych i czynników demograficznych na spożycie alkoholu w dni powszednie

W celu identyfikacji czynników wpływających na czas poświęcany na naukę przez uczniów przeprowadzono analizę regresji Poissona z studytime jako zmienną zależną. W modelu uwzględniono relacje rodzinne (famrel_good), liczbę nieobecności (absences), ilość wolnego czasu (freetime), płeć (sex) oraz poziom wykształcenia matki (Medu) i ojca (Fedu).

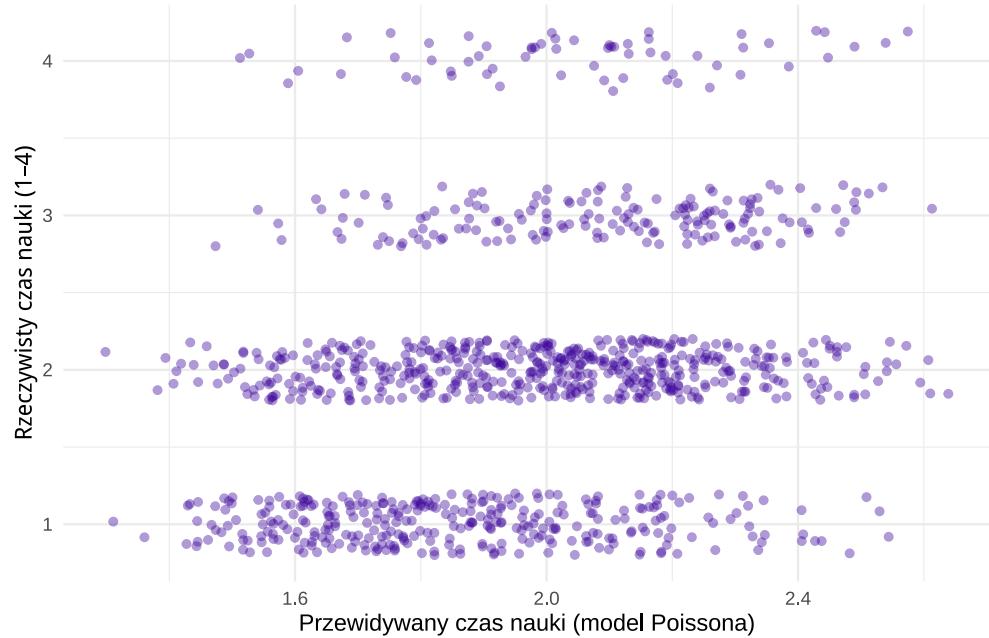
Spośród wszystkich analizowanych predyktorów istotne statystycznie okazały się tylko dwie zmienne: płeć oraz poziom wykształcenia matki. Dziewczęta, zgodnie z oczekiwaniemi, uczą się przeciętnie więcej niż chłopcy (sexM, $p < 0.001$), co może wynikać z różnic w podejściu do obowiązków szkolnych, samodyscypliny lub presji społecznej.

Z kolei wpływ wykształcenia matki (Medu, $p = 0.03$) potwierdza, że środowisko domowe, a w szczególności postawy edukacyjne matek, mogą istotnie oddziaływać na zaangażowanie dzieci w naukę. Brak istotnego wpływu wykształcenia ojca (Fedu) może sugerować, że to matki częściej pełnią rolę wzorca edukacyjnego i są bardziej zaangażowane w codzienny nadzór nad nauką dzieci – szczególnie w młodszym wieku.

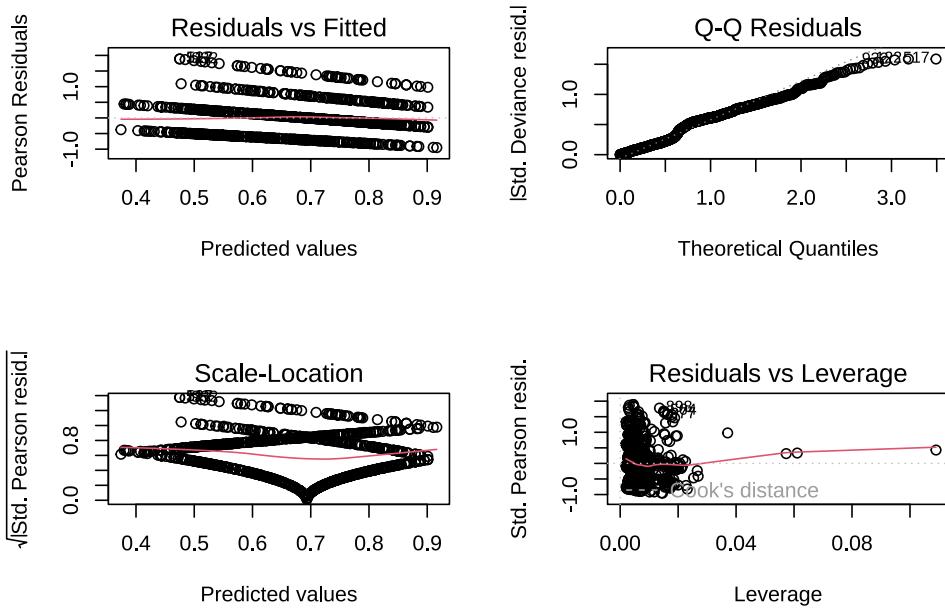
Pozostałe predyktory, takie jak relacje rodzinne, absencje czy wolny czas, nie osiągnęły istotności statystycznej, choć ich kierunki były zgodne z intuicyjnymi założeniami – np. większa liczba nieobecności wiązała się z mniejszą ilością deklarowanej nauki, a więcej wolnego czasu nie przekładało się wprost na większe zaangażowanie edukacyjne.

Podsumowując, model wskazuje, że czynniki społeczne i rodzinne – szczególnie płeć oraz postawa edukacyjna matek – mają istotny wpływ na czas poświęcany przez uczniów na naukę w tygodniu. Wyniki te mogą stanowić punkt wyjścia do dalszych analiz oraz działań wspierających rozwój postaw edukacyjnych w środowisku domowym.

Rzeczywisty vs przewidywany czas nauki uczniów



Wyniki modelu i diagnostyka



Ocena dopasowania modelu

Aby ocenić jakość dopasowania modelu Poissona przewidującego czas nauki uczniów, przeprowadzono standardową analizę diagnostyczną opartą na wykresach reszt.

Na wykresie Residuals vs Fitted obserwujemy systematyczne rozwarstwienie reszt wzduż prognozowanych wartości, co wskazuje na fakt, że model nie w pełni oddaje nieliniowości lub strukturę danych. Mimo braku wyraźnych odchyleń trendu, występują grupy punktów ułożonych warstwowo, co może być skutkiem dyskretnej skali zmiennej zależnej.

Wykres Q-Q Residuals pokazuje, że rozkład reszt istotnie odbiega od normalnego - szczególnie w wyższych kwantylach. To odchylenie może sugerować, że model nie do końca radzi sobie z przewidywaniem najwyższych poziomów czasu nauki i że warto rozważyć alternatywną specyfikację modelu lub uwzględnienie dodatkowych zmiennych.

Na wykresie Scale-Location nie widać wyraźnych oznak heteroskedastyczności - rozproszenie reszt jest stosunkowo równomierne, a linia trendu nie wykazuje silnego nachylenia. To świadczy o stabilności wariancji błędów w analizowanym zakresie przewidywanych wartości.

Wreszcie, wykres Residuals vs Leverage nie ujawnia obecności obserwacji ekstremalnie wpływowych – nie zaobserwowano punktów przekraczających próg dystansu Cooka. Model nie wydaje się być zdominowany przez pojedyncze przypadki.

Podsumowując: model wykrywa istotne zależności (np. wpływ płci czy wykształcenia matki), jednak jego dopasowanie nie jest idealne – szczególnie w wyższych poziomach studytime. Możliwe, że przyczyną są ograniczenia wynikające z uporządkowanej, niskopoziomowej skali czasu nauki. Pomimo tego, model dostarcza przydatnych wniosków o czynnikach wpływających na zaangażowanie uczniów w naukę.

Podsumowanie etapu 2: Relacje rodzinne a zachowania społeczne uczniów

Przeprowadzona analiza wykazała, że jakość relacji rodzinnych uczniów odgrywa zróżnicowaną rolę w kształtowaniu ich zachowań społecznych. Pozytywne więzi rodzinne nie wpływają bezpośrednio na częstotliwość wyjść z przyjaciółmi – kluczowymi czynnikami w tym przypadku okazały się wiek oraz ilość wolnego czasu. Oznacza to, że potrzeba kontaktów rówieśniczych realizowana jest głównie w oparciu o zasoby czasowe i etapy rozwoju, niezależnie od jakości środowiska domowego.

Zupełnie inny obraz wyłania się w przypadku spożycia alkoholu. Wyniki analiz jednoznacznie wskazują, że dobre relacje rodzinne pełnią rolę czynnika ochronnego przed nadmiernym spożywaniem alkoholu – zarówno w dni powszednie, jak i w weekendy. Uczniowie deklarujący pozytywne więzi rodzinne znacznie rzadziej sięgają po alkohol, szczególnie w ciągu tygodnia.

Co ciekawe, zaobserwowano, że wpływ relacji rodzinnych na spożycie alkoholu jest silniejszy w tygodniu niż w weekendy. Może to wynikać z różnic w kontekście społecznym picia – w dni robocze młodzież częściej spożywa alkohol w gronie znajomych, z którymi dzieli podobny rytm dnia (np. szkoła, praca), a nie z rodziną. W weekendy natomiast wspólne spotkania rodzinne, obiady, czy imprezy okolicznościowe mogą sprzyjać okazjonalnemu piciu także w obecności dorosłych członków rodziny. To tłumaczy, dlaczego silne relacje rodzinne nie eliminują całkowicie spożycia alkoholu w weekend, a raczej zmieniają jego kontekst z ukrytego i ryzykownego – na bardziej społeczny i kontrolowany.

Dodatkowo, wyższe ryzyko nadmiernego picia wiąże się z płciąorską, starszym wiekiem oraz większą aktywnością towarzyską – co pozostaje spójne z wcześniejszymi badaniami dotyczącymi zachowań ryzykownych wśród młodzieży.

Również w kontekście frekwencji szkolnej i czasochłonności nauki relacje rodzinne wykazują istotne zależności. Uczniowie deklarujący dobre relacje rodzinne znacznie rzadziej opuszczają zajęcia, co potwierdziły zarówno analizy rozkładów, jak i model regresji Poissona. Warto zauważać, że nieobecności są również istotnie powiązane z innymi zachowaniami ryzykownymi – takimi jak spożycie alkoholu – co sugeruje, że mogą one stanowić formę buntu lub ucieczki od presji edukacyjnej.

Jeśli chodzi o czas poświęcany na naukę, najważniejszym czynnikiem okazuje się płeć – dziewczęta przeciętnie uczą się więcej niż chłopcy. Ciekawym wynikiem jest też wpływ wykształcenia matki (Medu), który – w przeciwnieństwie do ojca – istotnie wpływa na zaangażowanie dziecka w naukę. Może to wynikać z większej obecności matek w życiu codziennym dzieci oraz z tego, że to one częściej kształtują edukacyjne nawyki swoich dzieci po szkole.

Podsumowując, wsparcie rodzinne i styl życia ucznia wpływają nie tylko na zachowania społeczne i używki, ale także na systematyczność i motywację edukacyjną. Wyniki te pokazują, że spójna i wspierająca atmosfera domowa pełni wielowymiarową funkcję ochronną – ograniczając zarówno ryzykowne zachowania, jak i poprawiając frekwencję oraz zaangażowanie w naukę.

Etap 3: Wpływ stylu życia na wyniki edukacyjne ucznia

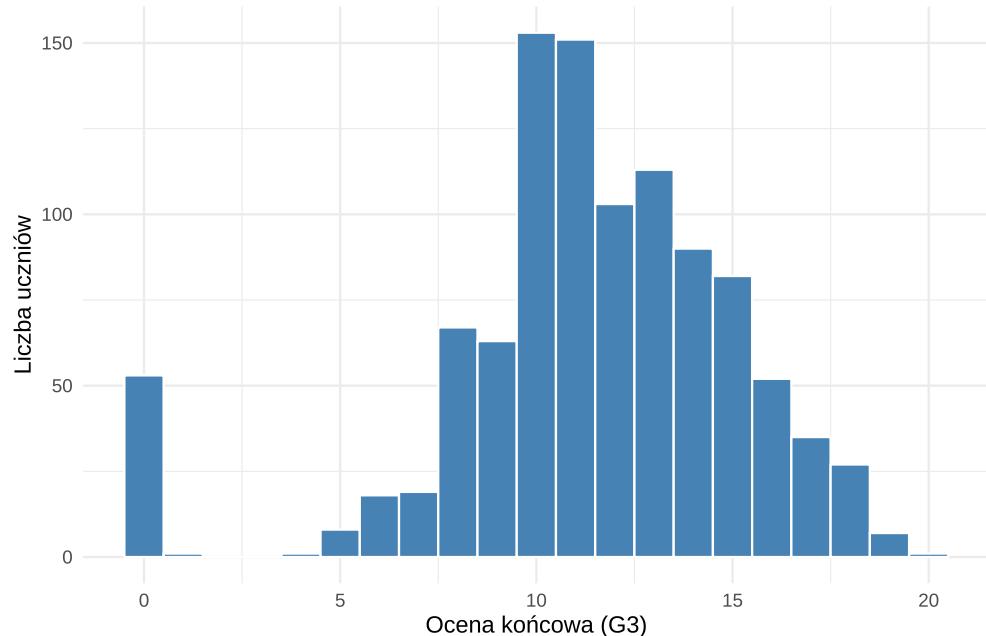
Kolejnym ogniwem analizowanego efektu domina jest styl życia ucznia, który może mieć istotne znaczenie dla jego wyników w nauce. Codzienne wybory młodzieży – takie jak sposób spędzania czasu wolnego, zaangażowanie w życie towarzyskie, obecność na zajęciach czy nawyki związane z użytkami – wpływają nie tylko na ich dobrostan, ale również na poziom koncentracji, motywację oraz systematyczność, czyli czynniki bezpośrednio przekładające się na osiągnięcia edukacyjne.

W tej części analizie poddano zależność pomiędzy różnymi aspektami stylu życia (takimi jak spożycie alkoholu, liczba nieobecności, liczba godzin nauki, czas wolny i częstotliwość spotkań z przyjaciółmi) a oceną końcową ucznia (G3). Celem jest określenie, które elementy codziennego funkcjonowania mają kluczowe znaczenie dla wyników szkolnych oraz w jakim stopniu mogą one pośredniczyć w działaniu wcześniejszych czynników, takich jak relacje rodzinne.

Analiza ta pozwala zidentyfikować czynniki ryzyka oraz elementy ochronne w kontekście edukacyjnym, a także wskazuje, które aspekty stylu życia powinny być wspierane lub korygowane w ramach działań profilaktycznych i wychowawczych.

Jaką ocenę ostatecznie dostali uczniowie?

Rozkład końcowej oceny G3



Analiza rozkładu końcowej oceny (G3) wskazuje, że większość uczniów osiąga wyniki w środkowym zakresie skali, najczęściej między 9 a 14 punktami. Rozkład jest lekko prawoskoniasty, co sugeruje, że oceny bardzo wysokie (bliskie maksymalnej wartości 20) należą do rzadkości. Uczniowie ogólnie radzą sobie przeciętnie lub dobrze - co odzwierciedla dominująca liczba ocen zbliżonych do środka skali.

Na uwagę zasługuje również wyraźny pik dla wartości G3 = 0. Tak duża liczba zerowych ocen wskazuje prawdopodobnie nie na rzeczywiste porażki edukacyjne, lecz raczej na nieukończenie kursu - np. z powodu nieobecności na egzaminie, rezygnacji z zajęć lub przyczyn formalnych.

Ze względu na ich odmienny charakter, obserwacje z G3 = 0 zostały wykluczone z dalszej analizy. Celem niniejszego projektu jest bowiem identyfikacja czynników wpływających na uzyskane wyniki edukacyjne, a nie na samo ryzyko rezygnacji z nauki. Uwzględnienie tej specyficznej grupy mogłoby zaburzyć modele predyktocyjne i wprowadzać trudne do interpretacji zniekształcenia.

Modelowanie statystyczne

W celu zbadania wpływu stylu życia na osiągnięcia edukacyjne ucznia przeprowadzono model regresji liniowej, w którym zmienną objaśnianą była końcowa ocena (G3). W modelu uwzględniono czynniki związane z codziennym funkcjonowaniem ucznia: spożycie alkoholu (Dalc, Walc), częstotliwość spotkań towarzyskich (goout), ilość wolnego czasu (freetime), czas przeznaczany na naukę (studytime), liczbę nieobecności (absences) oraz jakość relacji rodzinnych (famrel).

Model ten stanowił kluczowy etap weryfikacji hipotezy o efektach stylu życia na wyniki edukacyjne, uwzględniając przy tym istotne dla młodzieży czynniki ryzyka i wsparcia środowiskowego.

```
model_g3 <- glm(G3 ~ Dalc + Walc + goout + freetime +
                  studytime + absences + famrel,
                  data = students[students$G3 > 0, ])
summary(model_g3)

##
## Call:
## glm(formula = G3 ~ Dalc + Walc + goout + freetime + studytime +
##       absences + famrel, data = students[students$G3 > 0, ])
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.36097   0.56323  21.947 < 2e-16 ***
## Dalc        -0.24896   0.12439  -2.001  0.0456 *
## Walc        -0.10123   0.09493  -1.066  0.2866
## goout       -0.16908   0.08876  -1.905  0.0571 .
## freetime    -0.11442   0.09083  -1.260  0.2081
## studytime   0.49738   0.10882   4.570 5.48e-06 ***
## absences    -0.08494   0.01422  -5.973 3.26e-09 ***
## famrel       0.12800   0.09701   1.319  0.1873
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 7.652611)
##
## Null deviance: 8396.4 on 990 degrees of freedom
## Residual deviance: 7522.5 on 983 degrees of freedom
## AIC: 4839
##
## Number of Fisher Scoring iterations: 2
```

Wyniki modelu: wpływ zachowań i środowiska ucznia na ocenę końcową

Wyniki modelu regresji liniowej potwierdzają intuicyjne założenie, że jednym z najsilniejszych czynników wpływających na końcową ocenę ucznia (G3) jest czas poświęcony na naukę w tygodniu (studytime). Im więcej uczeń się uczy, tym wyższe osiąga wyniki - efekt ten jest silnie istotny statystycznie (studytime, $p < 0.001$), co potwierdza skuteczność systematycznej nauki jako strategii poprawy osiągnięć edukacyjnych.

Model ujawnił również istotny negatywny wpływ spożycia alkoholu w dni powszednie (Dalc, $p = 0.0456$) na ocenę końcową. Może to wskazywać na to, że picie alkoholu w środku tygodnia koliduje z codziennymi obowiązkami szkolnymi, obniżając koncentrację i motywację. Co ciekawe, spożycie alkoholu w weekend nie wykazało istotności statystycznej (Walc, $p = 0.88$) – być może dlatego, że weekendowe picie ma bardziej zaplanowany i społecznie akceptowalny charakter i nie koliduje bezpośrednio z obowiązkami szkolnymi.

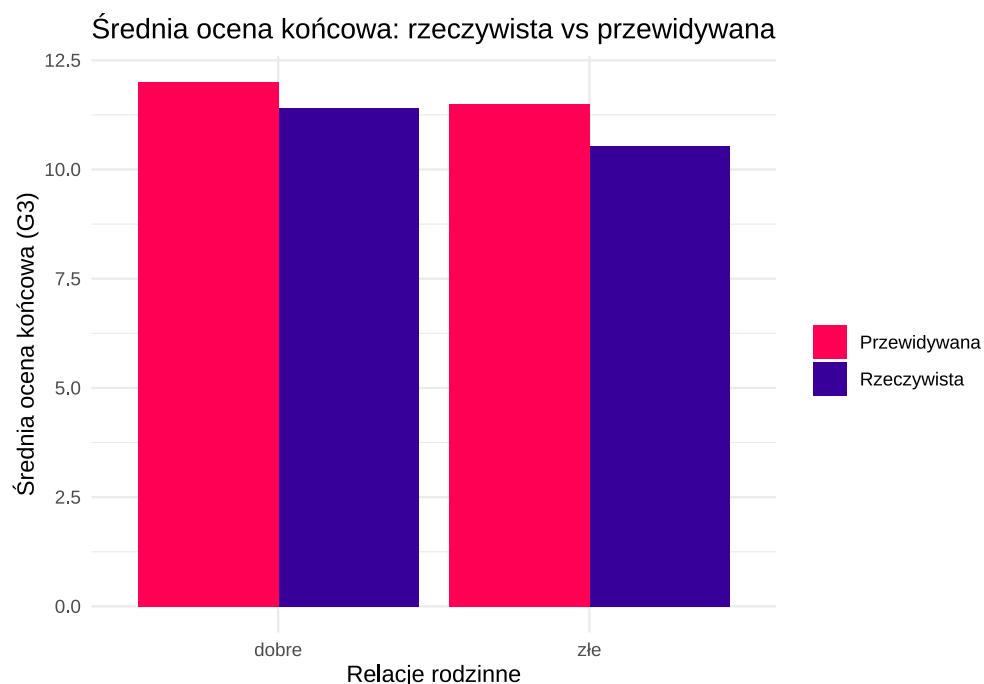
Wyniki wskazują również na marginalnie istotny negatywny wpływ liczby wyjść z przyjaciółmi (goout, $p = 0.0571$), co może sugerować, że intensywna aktywność towarzyska częściowo ogranicza czas przeznaczony na naukę lub regenerację. Zmienna freetime, opisująca ogólną ilość wolnego czasu ucznia, nie wykazała natomiast istotnego wpływu na ocenę końcową.

Warto jednak zauważyć, że w rzeczywistości freetime i goout mogą być ze sobą powiązane – dla wielu uczniów wolny czas w dużej mierze oznacza właśnie spotkania towarzyskie. Co więcej, wyjścia z przyjaciółmi często trwają dłużej niż planowano, a ich spontaniczny charakter może powodować, że realnie kolidują z nauką, nawet jeśli formalnie uczeń deklaruje posiadanie wolnych godzin. Z tego względu goout może silniej odzwierciedlać niekontrolowaną formę spędzania czasu niż sama liczba godzin wolnych (freetime).

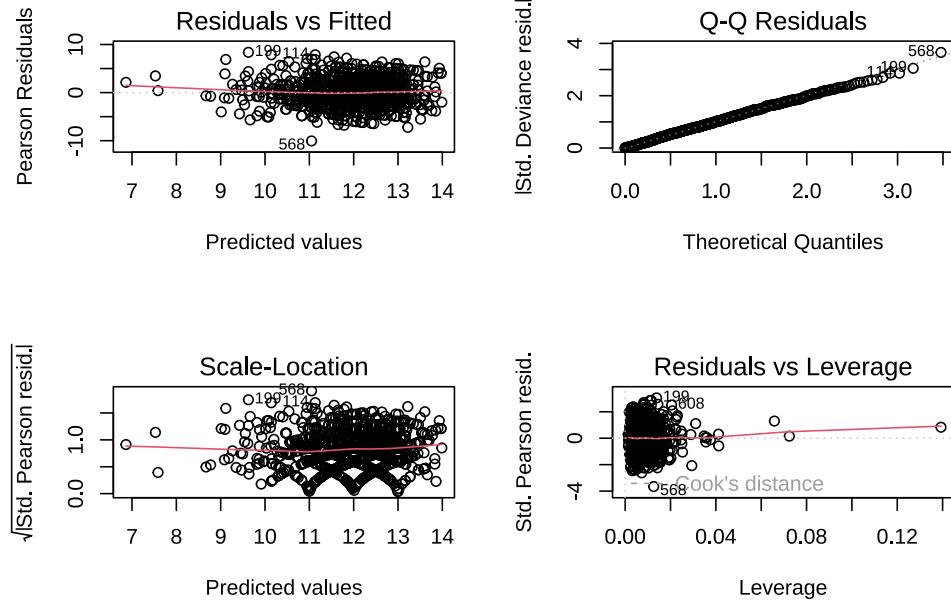
Na szczególną uwagę zasługuje zmienna absences – liczba nieobecności w szkole – która okazała się silnie istotna statystycznie (absences, $p < 0,001$) i wyraźnie negatywnie związana z wynikami końcowymi. Wynik ten wskazuje, że uczniowie nieobecni częściej osiągają niższe oceny, co potwierdza znaczenie systematycznej obecności i bieżącego uczestnictwa w zajęciach.

Zmienna famrel, opisująca jakość relacji rodzinnych, nie osiągnęła poziomu istotności statystycznej (famrel, $p = 0.1873$), co może oznaczać, że jej wpływ jest słabszy lub pośredni – np. przez styl życia ucznia, a nie bezpośrednio przez wynik końcowy.

Podsumowując, analiza wskazuje, że kluczową rolę w kształtowaniu wyników edukacyjnych odgrywa styl życia ucznia – zwłaszcza czas poświęcanego na naukę oraz regularna obecność w szkole. Inne czynniki, takie jak relacje społeczne i spożycie alkoholu w tygodniu, również mogą wpływać na wyniki, choć ich znaczenie jest umiarkowane. Równowaga między życiem prywatnym a obowiązkami szkolnymi wydaje się mieć istotne znaczenie dla sukcesów edukacyjnych młodzieży.



Wyniki modelu i diagnostyka



Ocena jakości dopasowania modelu

Aby ocenić jakość dopasowania modelu regresji liniowej dla zmiennej zależnej G3, przeprowadzono standardową analizę diagnostyczną.

Wykres *Residuals vs Fitted* nie wskazuje na obecność nieliniowości ani poważnych odchyлеń systematycznych - reszty są równomiernie rozproszone wokół osi zerowej, co sugeruje poprawne odwzorowanie zależności między zmiennymi. Nie widać wyraźnych skupień, lejków ani istotnych trendów, które mogłyby świadczyć o niedopasowaniu modelu.

Wykres *Q-Q Residuals* prezentuje reszty bardzo dobrze dopasowane do rozkładu normalnego. Punkty niemal idealnie układają się wzduż linii referencyjnej, z jedynie nieznacznymi odchyleniami na końcach. Oznacza to, że założenie normalności reszt jest w dużym stopniu spełnione.

Wykres *Scale-Location* nie ujawnia oznak heteroskedastyczności – wariancja reszt jest względnie stała w całym zakresie przewidywanych wartości, a linia trendu pozostaje praktycznie pozioma. Wzorzysty układ punktów najprawdopodobniej wynika z dyskretnej skali ocen (G3) i nie świadczy o błędzie modelowym.

Na wykresie *Residuals vs Leverage* widoczne są pojedyncze przypadki o podwyższonej dźwigni, jednak żaden z punktów nie przekracza wyraźnie granicy Cook's distance. Model nie jest zdominowany przez obserwacje odstające.

Wnioski:

Diagnostyka wskazuje, że model został dopasowany poprawnie. Nie stwierdzono poważnych naruszeń założeń regresji liniowej - rozkład reszt, nieliniowość, homoskedastyczność oraz brak dominujących obserwacji pozostają w akceptowalnych granicach. Model można uznać za wiarygodny zarówno do interpretacji współczynników, jak i do celów predykcyjnych.

Podsumowanie etapu 3: Wpływ stylu życia na wyniki edukacyjne ucznia

Podsumowując, analiza wskazuje, że styl życia ucznia wywiera istotny wpływ na jego wyniki edukacyjne. Najśilniejszym predyktorem sukcesu szkolnego okazał się czas poświęcany na naukę – uczniowie systematycznie uczący się w tygodniu osiągają wyraźnie wyższe oceny. Równie istotnym czynnikiem okazała się liczba nieobecności – osoby często opuszczające zajęcia mają znaczco gorsze wyniki, co podkreśla znaczenie obecności i aktywnego uczestnictwa w lekcjach.

Wśród czynników ryzyka wyraźnie zaznacza się spożycie alkoholu w dni powszednie, które negatywnie korreluje z ocenami – może to sugerować, że takie zachowania bezpośrednio zaburzają rytm nauki i obniżają zdolność do koncentracji. Dla kontrastu, picie w weekend nie miało istotnego wpływu na ocenę końcową, co może być efektem bardziej kontrolowanego, okazjonalnego charakteru takich sytuacji.

Zmienna opisująca aktywność towarzyską (goout) również wykazała marginalny negatywny wpływ na wyniki. Choć sama ilość wolnego czasu (freetime) nie była istotna statystycznie, należy zauważać, że realne wykorzystanie tego czasu – np. na wyjścia z przyjaciółmi – może kolidować z nauką. Szczególnie że tego typu aktywności często mają spontaniczny i trudny do przewidzenia charakter, przez co rzeczywisty wpływ na wyniki może być większy, niż wskazuje na to statystyka.

Co interesujące, jakość relacji rodzinnych (famrel) nie okazała się istotnym czynnikiem bezpośrednio wpływającym na ocenę końcową. Może to oznaczać, że jej rola jest bardziej pośrednia – np. poprzez kształtowanie stylu życia lub postaw względem szkoły, niż poprzez bezpośredni wpływ na osiągnięcia edukacyjne.

Wnioski te wzmacniają postulat, że skuteczne działania wychowawcze i profilaktyczne powinny koncentrować się nie tylko na wartościach rodzinnych, ale także na promowaniu systematyczności, obecności w szkole oraz ograniczaniu ryzykownych zachowań w codziennym funkcjonowaniu ucznia.

Etap 4: Aspiracje edukacyjne jako konsekwencja doświadczeń szkolnych i życiowych

Kolejnym ogniwem analizowanego efektu domina są aspiracje edukacyjne ucznia, czyli decyzja o kontynuowaniu nauki po szkole średniej. Jest to jeden z najważniejszych momentów przejściowych w życiu młodzieży, który może determinować dalszy rozwój osobisty, zawodowy i społeczny.

Wybór ten nie jest jednak oderwany od wcześniejszych doświadczeń - podejmowane decyzje edukacyjne kształtują się w kontekście relacji rodzinnych, stylu życia, codziennych nawyków i dotychczasowych osiągnięć. Uczniowie, którzy osiągają dobre wyniki, systematycznie się uczą, mają wsparcie środowiskowe i unikają ryzykownych zachowań, mogą częściej deklarować chęć dalszej edukacji.

W tej części analizie poddano, w jakim stopniu wyniki w nauce, styl życia oraz czynniki rodzinne i społeczne wpływają na prawdopodobieństwo kontynuowania nauki (zmienna higher). Pozwoli to zidentyfikować, które obszary wsparcia lub ryzyka mają kluczowe znaczenie dla decyzji edukacyjnych młodzieży.

Czy uczniowie chcą kontynuować naukę?



Wykres jednoznacznie pokazuje, że zdecydowana większość uczniów deklaruje chęć kontynuowania nauki po ukończeniu szkoły średniej. Tylko niewielka część respondentów zadeklarowała brak takich planów, co może świadczyć o wysokich aspiracjach edukacyjnych w badanej populacji.

Modelowanie statystyczne

W celu zbadania czynników wpływających na aspiracje edukacyjne uczniów przeprowadzono model regresji logistycznej, w którym zmienną objaśnianą była deklaracja chęci kontynuowania nauki po szkole średniej (higher). W modelu uwzględniono kluczowe aspekty stylu życia oraz funkcjonowania ucznia: spożycie alkoholu (Dalc, Walc), liczbę nieobecności (absences), czas nauki (studytime), jakość relacji rodzinnych (famrel), aktywność towarzyską (goout) oraz rzeczywisty wynik edukacyjny (G3).

Model ten stanowi istotny element weryfikacji hipotezy o długofalowym wpływie środowiska rodzinnego i stylu życia na postawy uczniów wobec edukacji. Pozwala zidentyfikować czynniki, które mogą wspierać lub organiczać ambicje młodzieży dotyczące dalszego kształcenia.

```

students$higher_bin <- ifelse(students$higher == "yes", 1, 0)

model_higher <- glm(higher_bin ~ G3 + Dalc + Walc +
                      goout + freetime + studytime + absences + famrel,
                      data = students, family = binomial)

summary(model_higher)

##
## Call:
## glm(formula = higher_bin ~ G3 + Dalc + Walc + goout + freetime +
##       studytime + absences + famrel, family = binomial, data = students)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.46179   0.74350 -0.621   0.5345
## G3          0.17099   0.02700  6.334 2.40e-10 ***
## Dalc        -0.18296   0.14236 -1.285   0.1987
## Walc         0.06420   0.12377  0.519   0.6040
## goout       -0.02787   0.11465 -0.243   0.8080
## freetime    -0.21289   0.12323 -1.728   0.0841 .
## studytime   0.93097   0.19457  4.785 1.71e-06 ***
## absences    -0.03119   0.01461 -2.135   0.0327 *
## famrel      0.16209   0.12080  1.342   0.1797
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 608.46 on 1043 degrees of freedom
## Residual deviance: 512.57 on 1035 degrees of freedom
## AIC: 530.57
##
## Number of Fisher Scoring iterations: 6

```

Wyniki modelu: styl życia i wyniki w nauce jako predyktory planów edukacyjnych ucznia

Wyniki modelu regresji logistycznej pozwalają określić, które czynniki wpływają na prawdopodobieństwo kontynuowania nauki przez uczniów po zakończeniu szkoły średniej. Zgodnie z oczekiwaniami, kluczową rolę odgrywa końcowa ocena (G3) oraz czas poświęcony na naukę w tygodniu (studytime). Obie zmienne są silnie istotne statystycznie ($p < 0.001$), co nie dziwi – osoby osiągające lepsze wyniki oraz systematycznie uczące się wykazują większą gotowość do dalszego kształcenia.

Istotny statystycznie (absences, $p = 0.0327$) okazał się również wpływ liczby nieobecności (absences). Choć efekt ten jest mniej wyraźny niż w przypadku G3 i studytime, wskazuje na to, że uczniowie opuszczający zajęcia już na obecnym etapie edukacji rzadziej deklarują chęć dalszego kształcenia – co jest spójne z intuicyjnym rozumieniem zaanga

Na granicy istotności statystycznej znajduje się zmieniona (freetime, $p = 0.0841$), która – co ciekawe – przyjmuje znak ujemny. Może to sugerować, że osoby dysponujące większą ilością wolnego czasu rzadziej planują dalszą edukację. Jest to zaskakujące, ponieważ można by oczekwać, że osoby z dużą ilością wolnego czasu będą bardziej skłonne do kontynuowania nauki. W praktyce może to jednak oznaczać, że niektórzy uczniowie nie poświęcają czasu na rozwój edukacyjny, ponieważ ich styl życia ukierunkowany jest na inne aktywności – np. pracę zarobkową, opiekę nad bliskimi lub inne zobowiązania, które konkurują z edukacją.

Pozostałe zmienne, takie jak spożycie alkoholu (Dalc, Walc), wyjścia z przyjaciółmi (goout) oraz jakość relacji rodzinnych (famrel), nie osiągnęły istotności statystycznej w tym modelu. Choć ich wpływ mógł być zauważalny na wcześniejszych etapach analizy, w kontekście decyzji o kontynuacji nauki nie odgrywają one kluczowej roli. Warto jednak pamiętać, że ich efekt może być pośredni – np. poprzez wpływ na styl życia, motywację lub wyniki w nauce.

```
hoslem.test(model_higher$y, fitted(model_higher), g=10)
```

Test zgodności modelu (Hosmer-Lemeshow)

```
##  
## Hosmer and Lemeshow goodness of fit (GOF) test  
##  
## data: model_higher$y, fitted(model_higher)  
## X-squared = 6.234, df = 8, p-value = 0.621
```

Ocena dopasowania modelu

Otrzymana wartość $p = 0.621$ jest znacznie większa od poziomu istotności 0.05. Oznacza to, że nie ma podstaw do odrzucenia hipotezy o zgodności modelu z danymi. Innymi słowy, model logistyczny przewidujący plany kontynuowania nauki przez uczniów został dobrze dopasowany do danych – przewidywane prawdopodobieństwa nie odbiegają systematycznie od rzeczywistych wartości zmiennej zależnej.

Model można uznać za wiarygodny w kontekście analizy wpływu rozważanych predyktorów na decyzję ucznia o kontynuowaniu nauki.

Podsumowanie

Chociaż początkowa hipoteza sugerowała prosty, łańcuchowy model wpływu środowiska rodzinnego na kolejne etapy życia młodego człowieka – od rodzinnego zaplecza, przez relacje i codzienne wybory, aż po wyniki edukacyjne i decyzje dotyczące dalszej nauki – analiza pokazuje, że rzeczywistość jest znacznie bardziej wielowątkowa i nie zawsze liniowa. Życie nie układa się w idealny efekt domina, ale w wielu miejscach ten schemat znajduje potwierdzenie, choć z pewnymi zastrzeżeniami.

Wyraźnie widać, że to właśnie stabilność rodzinnego zaplecza, a w szczególności wspólne zamieszkiwanie rodziców, przekłada się na lepszą jakość relacji rodzinnych. Liczebność rodziny czy samo wykształcenie rodziców okazały się mniej istotne dla budowania bliskości i wsparcia w domu – chociaż wśród predyktorów stylu życia (szczególnie systematyczności w nauce) to wykształcenie rodziców, a według naszych analiz zwłaszcza matki, miało zauważalny, pozytywny wpływ na zaangażowanie uczniów w naukę. Warto więc podkreślić, że wsparcie edukacyjne i postawa rodziców mogą przekładać się na motywację młodzieży, niekoniecznie poprzez atmosferę w domu, ale przez praktyczne wzorce i codzienną organizację życia.

Jakość relacji rodzinnych nie zawsze prowadzi wprost do lepszych ocen, ale ma wyraźny wpływ na ograniczenie ryzykownych zachowań – takich jak nadużywanie alkoholu w dni powszednie czy opuszczanie zajęć. Uczniowie z pozytywnymi więzami rodzinnymi rzadziej sięgają po alkohol i mają mniej nieobecności, a te aspekty stylu życia są już bezpośrednio powiązane z wynikami edukacyjnymi. Z analizy wynika, że największe znaczenie dla sukcesów szkolnych mają regularna obecność na lekcjach oraz czas faktyczne poświęcany na naukę – natomiast absencja i codzienne picie alkoholu wyraźnie pogarszają wyniki.

Dopiero w kolejnym kroku, na poziomie końcowych ocen, rozstrzyga się gotowość do dalszej edukacji: uczniowie z najlepszymi wynikami i ci najbardziej zaangażowani w naukę najczęściej planują kontynuację nauki po szkole średniej. Choć wydawałoby się, że duża ilość wolnego czasu mogłaby temu sprzyjać, w praktyce działa to odwrotnie – osoby z większą ilością wolnego czasu rzadziej wybierają ścieżkę dalszej edukacji, prawdopodobnie angażując się w inne aktywności poza szkołą.

Ostatecznie, choć nie wszystko układa się w jednoznaczny, prostoliniowy efekt domina, nasza analiza potwierdza, że środowisko rodzinne, styl życia i edukacyjne postawy rzeczywiście tworzą powiązany ze sobą system, w którym nawet pośrednie wpływy mogą mieć duże znaczenie. Schemat ten, choć niepozbawiony wyjątków i zawiłości, odzwierciedla realne mechanizmy kształtujące przyszłość młodych ludzi.