**Strain, Victimisation, and Youth Offending: A Quantitative Analysis Using the Millennium Cohort Study (7th Sweep, 2018)**

**Word count – 2499**

**Introduction**

This report examines the drivers of youth offending in England and Wales using data from the Millennium Cohort Study (MCS), a representative longitudinal data set that includes self-reported youth offending. The primary aim is to assess whether victimisation experiences are associated with offending using a quantitative, theory driven approach.

The central theoretical framework is General Strain Theory (GST) developed by Agnew (1992). He builds on Robert Merton’s (1938) strain theory which argues that complex social structures and lack of institutionalised means to achieve cultural goals leads to delinquency. In this context, the lack of opportunities for young people causes a reliance on offending. GST outlines three types of strain. Notably, negative stimuli like victimisation can trigger emotions such as anger, which may lead to offending as a coping response (Agnew, 1992;Ostrowsky, M., & Messner, S., 2005; Merton, 1938).

Comparatively, labelling theory as an alternative perspective will be used to complement GST. Broadly, labelling theory argues, when people are labelled negatively, their internalised identity becomes a master status, leading to more offending as a response mechanism. This highlights how multiple theoretical frameworks help explain complex issues such as youth offending (Motz et al., 2020; Wiley et al., 2013; Chenane et al., 2021; Bernburg et al., 2006).

Based on these frameworks the following hypothesis will be tested: “young people who experience higher levels of victimisation in the past year will report more offending”.

**Data and Sampling Procedures**

The Millennium Cohort Study (MCS), a large-scale nationally represented longitudinal study that follows 19,000 young people born within the United Kingdom between September 2000 and January 2002. I have used the seventh sweep of data which was collected in 2018 with respondents being on average 17 years old. The MCS used a stratified, cluster random sample which oversampled areas that were disadvantaged or had high ethnic minority populations. This allowed for a robust study of the effects of disadvantages of children which ensures that findings can be cautiously generalised to the broader population of young people in the UK. Additionally, its focus on social and behavioural outcomes makes it well-suited for studying youth offending.

**Descriptive Results**

**The dependent variable**

The dependent variable used in the analysis was a summated self-reported offending scale, constructed by combining all variables that capture whether respondents had engaged in offending in the previous year. These include carrying a knife, physical aggression (with or without a weapon), theft (person or shop) graffiti, criminal damage, hacking, burglary, and being a member of a street gang. Missingness was accounted for during this process to only account for complete responses.

This operationalisation is heavily backed in criminological research with studies using a variety of scales to capture the breadth of offending rather than single item measures (Home Office, 2019) (McAra, L., & McVie, S., 2010). Additionally, Sweeten (2012) argues using variety scales provides high reliability and validity to measure future delinquency. This provides significant support for my research report as we are aiming to provide implications into youth offending therefore when the findings are high in validity, our implications hold strong representative value.

Descriptive statistics were calculated to get an initial understanding of how offending behaviour is distributed across the sample. Measures of central tendency were used indicating, a mean of 0.4085, a median of 0, a standard deviation of 0.8010 and a response range of 0-8. The distribution of the variable is heavily skewed to the right, as shown in Figure 1 with the majority of respondents reporting no or very little offending. Presenting this descriptive overview is critical as it affirms our use of the variable in our regression analysis as it demonstrates the scales ability to capture variation in offending. However, it is crucial to consider limitations that arise during self-reported offending for example recall bias and underreporting. These issues may affect the reliability of the data and lead to an underestimation of true offending levels. Thus, having implications for the validity of the findings and the generalisability of policy recommendations.

A graph of a number of officers committed

AI-generated content may be incorrect.

Figure 1 - Distribution of Dependent Variable

**The independent variable**

The independent variable used in the analysis is a summated victimisation scale, constructed by combining four variables from the MCS: whether you have been a victim of threats, physical violence, weapon use, theft. This aims to capture all types of victimisation within the past year to be interpreted easier over single frequency measures (Home Office, 2019) (McAra, L., & McVie, S., 2010).

The victimisation scale was designed to reflect GST (Agnew, 1992), which posits that exposure to negative stimuli like threats or violence can generate emotional strain that increases the likelihood of offending. Prior empirical studies demonstrated that cumulative victimisation, especially with youth offenders was a key driver into the higher rates of delinquency (e.g., Home Office, 2019; Villadsen & Fitzsimons, 2019; Hay & Evans, 2006, Ostrowsky & Messner, 2005). Thus, they provide clarity into using a victimisation scale to capture accumulative strain experienced by youth, aligning both theoretically and empirically with the hypothesis of the research report. However, not all victimisation leads to offending, and individual responses may vary depending on other factors such as a lack of social support – none of which is captured in this measure.

A graph with numbers and a bar

AI-generated content may be incorrect.Descriptive statistics indicate a mean of 0.6337, median of 0, standard deviation of 0.8740, values ranging 0-4. The distribution is positively skewed indicating that the majority (50%) little or no victimisation. However, they indicate there is a minority experiencing victimisation which is empirical evidence for understanding how strain leads to offending (Agnew, 1992). To visualise this, Figure 2 confirms the use of treating the variable as discrete numerical scale in regression analysis.

Figure 2 - Distribution of Independent Variable

**Is my Independent Variable Associated with Youth Offending?**

**Initial test of hypothesis**

To test my hypothesis, a descriptive and inferential analysis was conducted. The preliminary descriptive analysis revealed a clear trend: as the victimisation increased, the mean of level of offending increased. For example, the average score of respondents receiving no victimisation reported an offending score of 0.19 whilst respondents receiving all four types of victimisation increase to 1.74. This suggests a significant association between victimisation and delinquent behaviour.

To formally assess this relationship, I conducted a bivariate linear regression with offending-scale as the dependent variable and victimisation-scale as the independent variable. The results indicated a statistically significant relationship between victimisation and self-reported offending (β = 0.382, p < 0.001; See Appendix B) Additionally, the model explains 17.4% of the variation (R^2 = 0.174) which suggests a moderate explanation for the associated relationship. This is significant as the results directly align with General Strain Theory, which argue exposure to negative stimuli, such as victimisation, can produce emotional strain and lead to higher rates of offending. Further, these results align with empirical studies such as Home Office (2019) who concluded, younger victims were more likely to offend.

However, while the effect size (R^2) is substantial enough to be of theoretical relevance, explaining 17.4% suggests the importance of considering other factors such as peer influence, which this regression fails to measure. Moreover, whilst the MCS is a representative survey, this analysis draws on only one of the sweeps of data collection. This can render the findings effectively cross-sectional and potentially limit our ability to infer causality. Nonetheless, our findings are consistent with empirical studies (Home Office, 2019; Villadsen & Fitzsimons, 2021) and provide robust support for our theory-driven hypothesis.

**Adding Control Variables**

To build upon my bivariate analysis I have decided to include two control variables which should account for more variance. The first variable is gender, constructed as 1 = male, 2 = female/other. Statistics indicated a disparity between men and women with a 96% prison population of men, 84% of arrests were men. (Justice, 2025) Additionally, empirical studies consistently found this disparity, highlighting the need to understand the differences in offending (Svensson et al, 2017; McAra & Smith, 2004). This allows me to test if these disparities exist among young offenders. Moreover, gender as a binary variable may oversimplify the complex ways that gender identity and socialisation interact with offending. Future research may benefit from more nuanced measures that account for interactions between genders and incorporating intersectional analysis (Collins, 2000).

The second control variable is police contact, constructed by combining whether you have been stopped by the police or arrested in the past year thereby capturing a broader scope of justice system interaction. The operationalisation is directly informed by labelling theory, as well as prior studies which have supported this, concluding that contact with the justice system – especially at a young age - promotes delinquency (Beardslee et all, 2019). Morrow et al (2018) found that negative experiences arisen from contact from the police can form a non-compliant demeanour. While this variable captures a key element in criminological theory, it is limited as it does not differentiate between the type and/or quality of police contact which may have significant differing effects.

To further analyse, a multiple linear regression model was developed by incorporating these two variables. The model produced a statistically significant improvement in explained variance compared to the bivariate analysis. (R^2 = 0.236 – 0.174; See Appendix D for full regression table) This suggests the additional variables meaningfully contributed to the understanding of self-reported behaviour. Victimisation remained a strong predictor (β = 0.331, p < 0.001), though some of its effect size was reduced suggesting a potential overlap with police contact and gender. Male respondents had significantly higher offending scores than females/other respondents (β = 0.470, p < 0.001) which is consistent with previous literature and national statistics (Justice, 2025). Police contact was the strongest predictor in the model (β = 0.470, p < 0.001), providing strong empirical support for labelling theory. This finding reinforces existing studies that conclude contact with the justice system particularly through adolescence can amplify offending through stigma (Beardslee et al., 2019; Morrow et al., 2018).

Model assumptions, including linearity, normality and homoscedasticity were assessed using standard diagnostic plots. (See Appendix E) Whilst most assumptions were acceptable, the scale-location plot indicated mild heteroscedasticity, and the residuals vs fitted plot suggested some non-linearity. These issues may reduce the precision of estimates at higher predicted values, meaning we must account for their implications when advising on policy. Moreover, while the model explains 23.6% of variance, 75% is still unaccounted for. This indicates the need to incorporate additional contextual or structural factors that may be influencing youth offending which are not captured in this data set. Future research may expand, including variables such as peer influence, school experience.

**Is my Independent Variable Associated with Violent Youth Offending?**

To test whether higher levels of victimisation are also associated with violent offending, a new binary variable was created combining, carrying a knife or weapon, hitting or physically assaulting someone, and hitting someone with a weapon. The operationalisation is backed by the CPS definition (The Crown Prosecution Service, 2022) as well as supporting the earlier GST-based analysis, suggesting that more intense victimisation may escalate not just general delinquency but also serious violence. (Agnew, 1992) Several empirical studies also treat violent offending in this way. (Home Office, 2019; Villadsen & Fitzsimons, 2021; Sweeten, 2012). A new binary variable was created where 1 indicated engagement in any of these in the past year and 0 indicates none.

The resulting variable indicated 76% of respondents did not engage in violent offending whereas, 24% did. This reflects wider trends in youth crime statistics that most young people do not engage in violent acts, but a small percentage do exhibit these behaviours. (Youth Endowment Fund, 2023) Figure 3 shows the distribution differences.

A logistic regression model was used to estimate the probability of violent offending based on the victimisation scale whilst including control variables, police-contact and gender. All predictors were statistically significant (p < 0.001). Victimisation had the strongest association, with an odds ratio of 2.70 (95% CI: 2.48-2.93; See Appendix G, H), suggesting that each additional victimisation experience increased the odds of violent offending by 170%. Figure 4 visualises this in a predictive probabilities plot, illustrating how the likelihood of violent offending rises sharply with each unit increase in victimisation.

The control variables also had strong associations. Males increased the odds of violent offending by an odds ratio of 2.34 (134%), which is consistent with well-established criminological research and national statistics. Police contact had a strong effect (OR = 2.33), which offers empirical support for labelling theory, which argue primary interaction with the criminal justice system can reinforce secondary deviant identities, which may increase violent offending. (Beardslee et al, 2019; Morrow et al, 2018) However, we must consider that labelling does not always increase violent offending, particularly as this is a sweep of data collection, another sweep may present different findings, meaning we have to cautiously infer findings to apply to a wider demographic.

Model diagnostics confirmed several acceptable assumptions. Multicollinearity was ruled out due to the VIF values < 1.02. A confusion matrix was used to assess the models performance. The findings revealed: accuracy = 80.3%, precision = 66.5%, sensitivity = 38.4%.(See Appendix I, J,K) The accuracy and precision indicate a strong identification of true positives however, the low sensitivity suggests the model underpredicts violent offenders which may be due to a class imbalance due to the majority of non-violent responses. This limitation highlights the potential weakness of logistic regressions in imbalanced data, as high accuracy can mask the poor detection of minority classes. This raises concerns over the effectiveness for identifying high-risk youth in policy contexts.

Together, these findings extend our earlier analysis by showing that victimisation is not only associated with general offending but also significantly, increase the likelihood of serious violence among youth offenders, directly reflecting our central hypothesis.

A graph with a red line

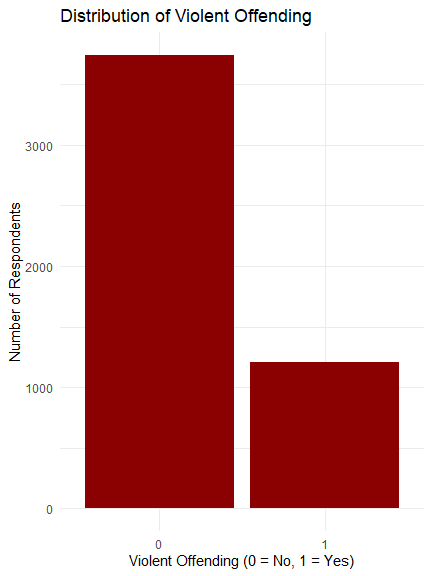
AI-generated content may be incorrect.

Figure 3 - Distribution of Violent Offending

Figure 4 – Predictive Probabilities Plot of Violent Offending by Victimisation Level

**Conclusion**

This report examined the relationship between victimisation and youth offending, drawing on Strain Theory and briefly, labelling theory. Using the Millennium Cohort Study, general and violent forms of offending were analysed. The findings provided strong empirical support for Strain Theory. Victimisation consistently predicted both general and violent offending, supporting GST’s core claim that strain from negative experiences can lead to delinquency. Incorporating control variables strengthened the models. Both police contact and gender were strongly associated with offending. This offered support for the alternative perspective, labelling theory, that frequent contact with the justice system does affect young people. The findings and previous literature on gender confirmed the positive association that demographic differences are present. Overall, this report highlights the importance of addressing victimisation and justice system contact in early intervention strategies. Whilst the models explain a meaningful portion of offending behaviour, we must consider that our research only explained a small proportion of variation. Future research should examine peer influence, family dynamics and structural inequalities to capture a more comprehensive understanding of youth behaviour. Finally, though self-reported data does provide rich insight, the presence of underreporting bias should be acknowledged as policy insights may hold less representative value if they underrepresent the impact on youth offenders.

**References**

Agnew, R. (1992) ‘Foundation for a general strain theory of crime and delinquency’, *Criminology*, 30(1), pp. 47-88. Available at: <https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1745-9125.1992.tb01093.x?saml_referrer> [Accessed 30 Apr. 2025].

Beardslee, J., Docherty, M., Mulvey, E., Schubert, C. and Piquero, A. (2019) ‘Does contact with the justice system deter or promote future delinquency?’, *Journal of Adolescent Health*, 65(5), pp. 682-688. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7317788/> [Accessed 30 Apr. 2025].

Bernburg, J.G., Krohn, M.D. and Rivera, C.J. (2006) ‘Official labelling, criminal embeddedness, and subsequent delinquency: A longitudinal test of labelling theory’, *Journal of Research in Crime and Delinquency*, 43(1), pp. 67-88.

Chenane, J.L., Veeh, C.A., Bratina, M.P. and Wooldredge, J.D. (2021) ‘Mapping the theoretical pathways from police contact to criminal activity for youth’, *Journal of Interpersonal Violence*. Available at: https://www.tandfonline.com/doi/full/10.1177/08862605211039328 [Accessed 30 Apr. 2025].

Collins, P.H. (2000) *Patricia Hill Collins: Intersecting oppressions*. [online] SAGE Publications. Available at: <https://us.sagepub.com/sites/default/files/upm-binaries/13299_Chapter_16_Web_Byte_Patricia_Hill_Collins.pdf> [Accessed 22 Apr. 2025].

Crown Prosecution Service (2022) *Violent crime*. Available at: <https://www.cps.gov.uk/crime-info/violent-crime> [Accessed 30 Apr. 2025].

Home Office (2019) *An analysis of indicators of serious violence: Findings from the Millennium Cohort Study and the Environmental Risk (E-Risk) Longitudinal Twin Study*. [online] London: Home Office. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\_data/file/849200/analysis-of-indicators-of-serious-violence.pdf [Accessed 30 Apr. 2025].

McAra, L. and McVie, S. (2010) ‘Youth crime and justice: Key messages from the Edinburgh Study of Youth Transitions and Crime’, *Criminology and Criminal Justice*, 10(2), pp. 179-209.

McAra, L. and Smith, D.J. (2004) *Gender and youth offending*. [online] University of Edinburgh, Centre for Law and Society. Available at: <https://www.research.ed.ac.uk/files/18766122/McAra_Smith_Gender_and_Youth_Offending.pdf> [Accessed 30 Apr. 2025].

Merton, R.K. (1938) ‘Social structure and anomie’, *American Sociological Review*, 3(5), pp. 672-682. Available at: <https://www.jstor.org/stable/2084686> [Accessed 30 Apr. 2025].

Ministry of Justice (2025) *Statistics on women and the criminal justice system 2023 (HTML)*. [online] GOV.UK. Available at: <https://www.gov.uk/government/statistics/women-and-the-criminal-justice-system-2023/statistics-on-women-and-the-criminal-justice-system-2023-html> [Accessed 19 Apr. 2025].

Morrow, W.J., Leone, M.C. and Piquero, A.R. (2018) *Examining the situational and suspect-level predictors of police use of force among a juvenile arrestee population*. [online] Centre on Juvenile and Criminal Justice. Available at: <https://www.cjcj.org/media/import/documents/examining_the_situational_and_suspect_level_predictors_of_police_use_of_force_among_a_juvenile_arrestee_population.pdf> [Accessed 30 Apr. 2025].

Motz, R.T., Nedelec, J.L. and Beaver, K.M. (2020) ‘Labelling effects of first juvenile arrests: Secondary deviance and secondary sanctioning’, *Youth Violence and Juvenile Justice*, 18(3), pp. 257-275.

Ostrowsky, M.K. and Messner, S.F. (2005) ‘Explaining crime for a young adult population: An application of general strain theory’, *Journal of Criminal Justice*, 33(5), pp. 464-476.

Sweeten, G. (2012) ‘Scaling criminal offending’, *Journal of Quantitative Criminology*, 28(3), pp. 533-557.

Svensson, R., Pauwels, L. and Weerman, F.M. (2017) ‘An examination of the interaction between morality and self-control in offending: A study of gender differences’, *Criminal Behaviour and Mental Health*, 27(6), pp. 565-575. Available at: <https://onlinelibrary.wiley.com/doi/full/10.1002/cbm.2065> [Accessed 30 Apr. 2025].

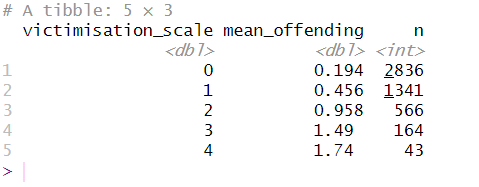
Villadsen, A. and Fitzsimons, E. (2021) *Adolescent victimisation, risky behaviours and police contact in the UK*. [online] Centre for Longitudinal Studies, Briefing Paper 13. Available at: <https://cls.ucl.ac.uk/wp-content/uploads/2021/04/CLS-briefing-paper-13-April-2021.pdf> [Accessed 30 Apr. 2025].

Wiley, S.A., Slocum, L.A. and Esbensen, F.-A. (2013) ‘The unintended consequences of being stopped or arrested: An exploration of the labelling mechanisms through which police contact leads to subsequent delinquency’, *Criminology*, 51(4), pp. 927-966.

Youth Endowment Fund (2023) *Children, violence and vulnerability 2023*. [online] Available at: <https://youthendowmentfund.org.uk/reports/children-violence-and-vulnerability-2023/> [Accessed 30 Apr. 2025].

**Appendix**

**A – Means table of bivariate analysis – shows victimisation scale relationship with offending**

****

**B – Bivariate linear regression results**

**A screenshot of a computer error

AI-generated content may be incorrect.**

**A group of graphs showing different values

AI-generated content may be incorrect.C – Bivariate linear regression diagnostic plots confirming model assumptions**

**D – Multiple level linear regression results**

**A screenshot of a computer

AI-generated content may be incorrect.**

**A group of graphs with numbers and lines

AI-generated content may be incorrect.E – Diagnostic plots, multiple linear regression. Confirming assumptions**

**F – Proportion table of violent offending**

**A number with numbers and symbols

AI-generated content may be incorrect.**

**G – Logistic regression results (violent offending)**

**A screenshot of a computer code

AI-generated content may be incorrect.**

**H – Odds ratios and Confidence Intervals**

**A number on a white background

AI-generated content may be incorrect.**

**I – VIF Results**

****

**J – Confusion Matrix**

**A number on a white background

AI-generated content may be incorrect.**

**K – Accuracy, Sensitivity, Precision Results**

**A screenshot of a computer code

AI-generated content may be incorrect.**

**R CODE**

rm(list = ls())

setwd("C:/R\_AST")

getwd()

#### Modelling Criminological Data Final Coursework ####

library(haven)

crime <- readRDS("mcs\_17.RDS")

#dependent variable #1 = alloffending (a combined scale of all different offending)

# 1. Create the self-reported offending scale (DV1)

crime$offending\_scale <- rowSums(crime[, c(

"carrying\_knife",

"physical\_aggression\_hit",

"physical\_aggression\_weapon",

"theft\_taken",

"theft\_stolen",

"graffiti",

"damage",

"hacking",

"hacking\_virus",

"burglary"

)], na.rm = TRUE)

# 2. Summary statistics

summary(crime$offending\_scale)

# 3. Measures of central tendency and dispersion

mean\_offending <- mean(crime$offending\_scale, na.rm = TRUE)

median\_offending <- median(crime$offending\_scale, na.rm = TRUE)

sd\_offending <- sd(crime$offending\_scale, na.rm = TRUE)

range\_offending <- range(crime$offending\_scale, na.rm = TRUE)

# 4. Print the stats

cat("Mean:", mean\_offending, "\n")

cat("Median:", median\_offending, "\n")

cat("Standard Deviation:", sd\_offending, "\n")

cat("Range:", range\_offending[1], "to", range\_offending[2], "\n")

# 5. Visualise the distribution

hist(crime$offending\_scale,

breaks = 10,

main = "Distribution of Self-Reported Offending Scale",

xlab = "Number of Offences Committed",

col = "lightblue",

border = "white")

#Independent variable = victimisation variable (combination of all victimisation) - directly referencing Agnew(1992)

# 1. Create the summated victimisation scale (IV)

crime$victimisation\_scale <- rowSums(crime[, c(

"victim\_threatened",

"victim\_violent",

"victim\_weapon",

"victim\_stolen"

)], na.rm = TRUE)

# 2. Summary of the victimisation scale

summary(crime$victimisation\_scale)

# 3. Central tendency and dispersion

mean\_victim <- mean(crime$victimisation\_scale, na.rm = TRUE)

median\_victim <- median(crime$victimisation\_scale, na.rm = TRUE)

sd\_victim <- sd(crime$victimisation\_scale, na.rm = TRUE)

range\_victim <- range(crime$victimisation\_scale, na.rm = TRUE)

# 4. Print the results

cat("Mean victimisation score:", mean\_victim, "\n")

cat("Median:", median\_victim, "\n")

cat("Standard Deviation:", sd\_victim, "\n")

cat("Range:", range\_victim[1], "to", range\_victim[2], "\n")

# 5. Histogram to visualise the distribution

hist(crime$victimisation\_scale,

breaks = 5,

main = "Distribution of Victimisation Scale",

xlab = "Number of Victimisation Types Experienced",

col = "tomato",

border = "white")

#Testing hypothesis using appropriate descriptive and inferential statistics.

#Is there an association between the independent and dependent variable

# Load required libraries

library(dplyr)

library(ggplot2)

library(effsize)

# 1. Descriptive: Mean offending by level of victimisation

crime %>%

group\_by(victimisation\_scale) %>%

summarise(

mean\_offending = mean(offending\_scale, na.rm = TRUE),

n = n()

)

# 2. Bivariate Linear Regression (DV = offending, IV = victimisation)

model\_simple <- lm(offending\_scale ~ victimisation\_scale, data = crime)

summary(model\_simple) # includes coefficient, p-value, and R-squared

# 3. R² - Model fit (variance explained)

r2 <- summary(model\_simple)$r.squared

cat("R² =", r2, "\n")

# Set up plotting area for 4 diagnostic plots

par(mfrow = c(2, 2))

plot(model\_simple)

#Adding control variables and performing multiple level linear regression

# Create police\_contact: 1 if either police.stopped or police.arrested is 1

crime$police\_contact <- ifelse(crime$police.stopped == 1 | crime$police.arrested == 1, 1, 0)

# Multiple regression model with IV + 2 control variables

model\_multiple <- lm(offending\_scale ~ victimisation\_scale + male + police\_contact, data = crime)

summary(model\_multiple)

# Display all four regression diagnostic plots

par(mfrow = c(2, 2)) # 2x2 grid layout

plot(model\_multiple) # Produces residual plots for assumption checks

#dependent variable 2 - Violent offending

# Create binary variable: 1 = engaged in violent offending, 0 = not

crime$violent\_offending <- ifelse(

crime$carrying\_knife == 1 |

crime$physical\_aggression\_hit == 1 |

crime$physical\_aggression\_weapon == 1,

1, 0

)

# Frequency and proportions

table(crime$violent\_offending)

prop.table(table(crime$violent\_offending))

# Visualisation

library(ggplot2)

ggplot(crime, aes(x = factor(violent\_offending))) +

geom\_bar(fill = "darkred") +

labs(title = "Distribution of Violent Offending",

x = "Violent Offending (0 = No, 1 = Yes)",

y = "Number of Respondents") +

theme\_minimal()

# Logistic regression (binary outcome)

model\_logistic <- glm(violent\_offending ~ victimisation\_scale + male + police\_contact,

data = crime, family = binomial)

summary(model\_logistic)

# Odds Ratios and 95% Confidence Intervals

exp(cbind(OR = coef(model\_logistic), confint(model\_logistic)))

#coefficient plot

library(sjPlot)

plot\_model(model\_logistic,

type = "est",

show.values = TRUE,

value.offset = 0.2,

title = "Logistic Regression Coefficients (Violent Offending)",

axis.labels = c("Police Contact", "Male", "Victimisation Scale"),

vline.color = "grey40")

# Effects plot (predictive margins / predicted probabilities)

library(sjPlot)

plot\_model(model\_logistic,

type = "pred",

terms = "victimisation\_scale", # You can do others too

title = "Predicted Probability of Violent Offending by Victimisation Level",

axis.title = c("Victimisation Score", "Predicted Probability"))

#Multi-collinearity

library(car)

vif(model\_logistic)

# Predict probabilities

crime$predicted\_prob <- predict(model\_logistic, type = "response")

# Convert to binary: threshold of 0.5

crime$predicted\_class <- ifelse(crime$predicted\_prob > 0.5, 1, 0)

# Confusion matrix

conf\_matrix <- table(Predicted = crime$predicted\_class, Actual = crime$violent\_offending)

print(conf\_matrix)

# Accuracy

accuracy <- mean(crime$predicted\_class == crime$violent\_offending)

cat("Accuracy:", round(accuracy, 3), "\n")

# Sensitivity (True Positive Rate)

TP <- conf\_matrix[2,2]

FN <- conf\_matrix[1,2]

sensitivity <- TP / (TP + FN)

cat("Sensitivity (Recall):", round(sensitivity, 3), "\n")

# Precision (Positive Predictive Value)

FP <- conf\_matrix[2,1]

precision <- TP / (TP + FP)

cat("Precision:", round(precision, 3), "\n")