Factors Predicting Problematic Phone Use

Siyi Zhu

May 2025

Contents

1 Introduction

This report analyzes the factors which predict problematic phone use. The dataset was extracted from ZPID's PsychArchives, on a research involving 122 adult participants, measuring scores on a personality measure (e.g. BIS-15), a psychological distress scale (HADS), and their problem phone use (MPPUS).

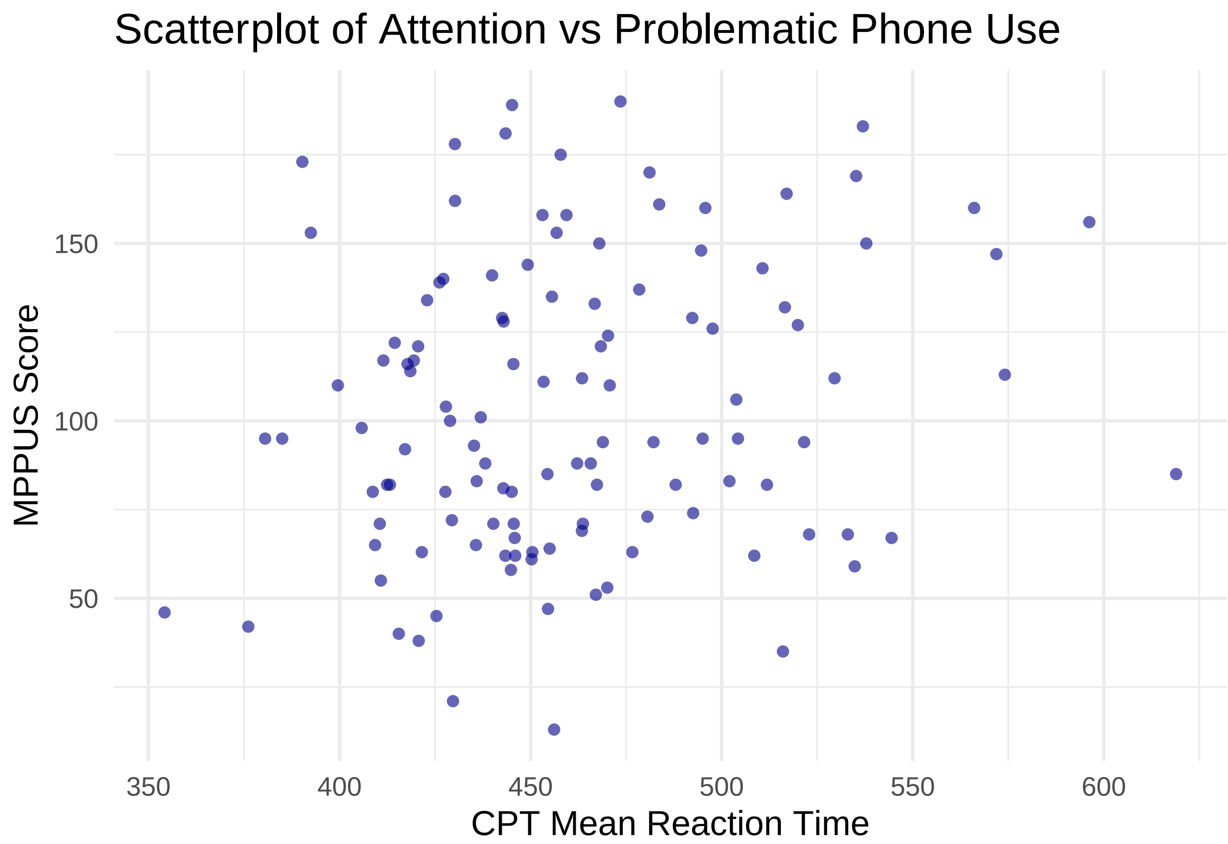
2 Exploratory Data Analysis (EDA)

Note that this report focuses on *which factors predict problematic phone use* rather than whom would be impacted, thus exploring the sample groups of participants that are at high risk of problematic phone use is not a part of this research.

2.1 Scatterplot – Attention vs Problematic Use

In exploring the relationship between attentional performance and problematic smartphone use, a scatterplot is shown with CPT Mean Reaction Time versus MPPUS Score. The results suggest a positive trend, where individuals with slower reaction times tend to report higher MPPUS scores. One observation was excluded due to missing or non-finite values.

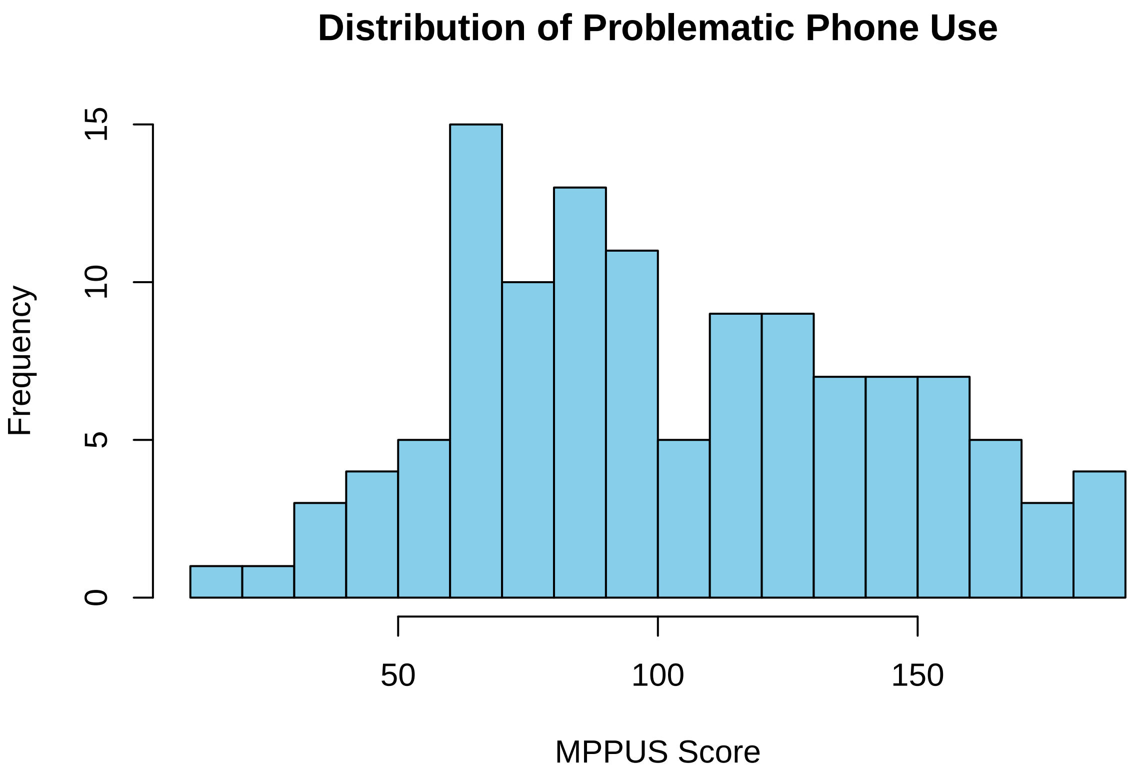
```
library(tidyverse)
library(ggplot2)
library(broom)
library(pROC)
library(car)
library(glmnet)
knitr::opts_chunk$set(echo = TRUE, fig.align = "center")

setwd("/Users/siyizhu/Desktop/stat3622/project")
df <- read_csv("Psychology of Smartphone Use Main Data Pluck et al.csv")
df_lm <- df %>%
  select(MPPUS__Total, CPTMeanTotal, GPA, Age, BIS15_Total, HADS_Total,
         Instagram, Facebook, Whatsapp) %>%
  drop_na()

ggplot(df_lm, aes(x = CPTMeanTotal, y = MPPUS__Total)) +
  geom_point(color = "darkblue", alpha = 0.6) +
  labs(
    title = "Scatterplot of Attention vs Problematic Phone Use",
    x = "CPT Mean Reaction Time",
    y = "MPPUS Score"
  ) +
  theme_minimal(base_size = 13)
```



2.2 Histogram of MPPUS Score

A histogram of MPPUS scores appears right-skewed, suggesting that extreme phone use behaviour is less common.

```
hist(df_lm$MPPUS__Total,
     breaks = 20,
     col = "skyblue",
     main = "Distribution of Problematic Phone Use",
     xlab = "MPPUS Score")
```



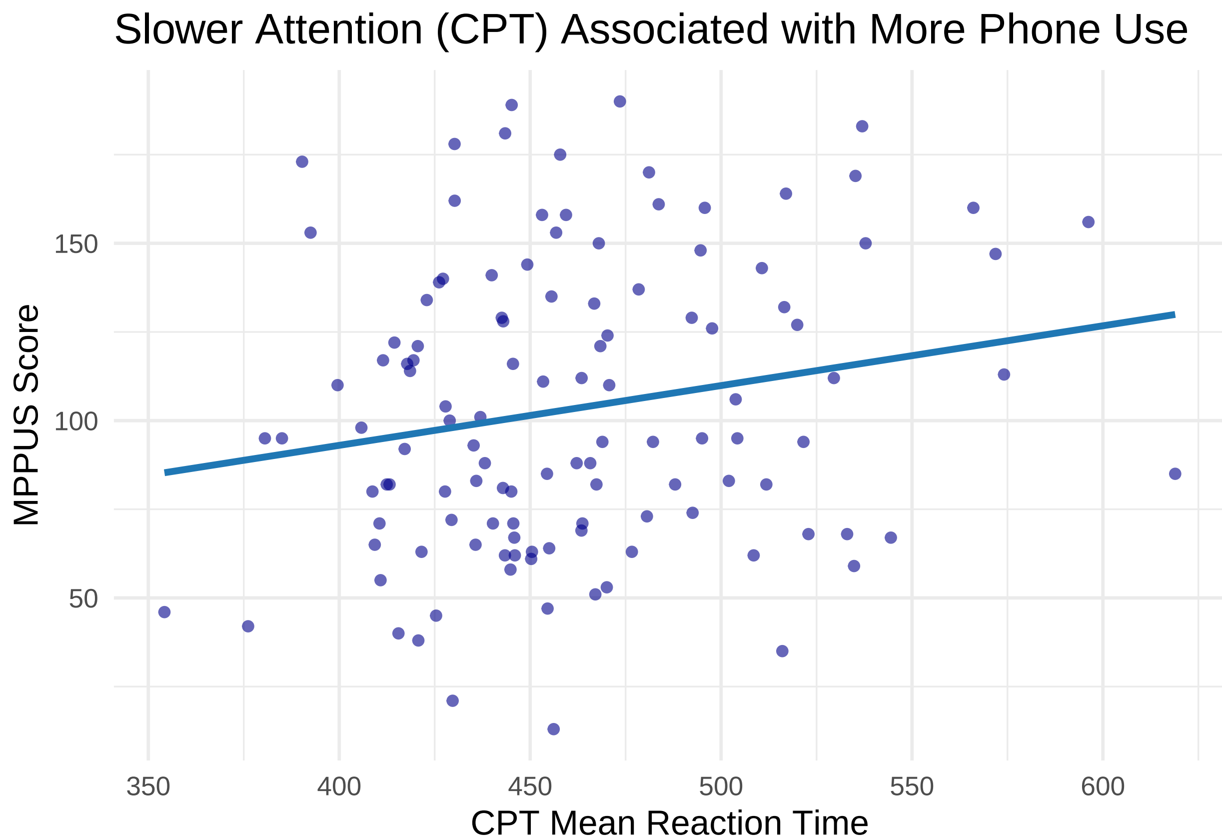
3 Visualisation

3.1 Linear Model

A multivariate linear regression indicates that as CPT reaction time increases, MPPUS also tends to increase. The effect size is modest.

```r
ggplot(df_lm, aes(x = CPTMeanTotal, y = MPPUS__Total)) +
  geom_point(color = "darkblue", alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE, color = "#1f77b4", linewidth = 1.2) +
  labs(
    title = "Slower Attention (CPT) Associated with More Phone Use",
    x = "CPT Mean Reaction Time",
    y = "MPPUS Score"
  ) +
  theme_minimal(base_size = 13)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



4 Analysis

4.1 Multicollinearity Check

It is also inituitive to check if multicollinearity is a concern as there are more than one psychological scales involved. In the following VIF output, it is evident that multicollinearity is not a serious concern in the model. All values are below 5.

```
model_lm <- lm(MPPUS__Total ~ CPTMeanTotal + GPA + Age + BIS15_Total +
                 HADS_Total + Instagram + Facebook + Whatsapp,
               data = df_lm)
vif(model_lm)
```

```
## CPTMeanTotal          GPA          Age  BIS15_Total   HADS_Total    Instagram
##     1.100823     1.064828     1.076434     1.346465     1.240563     2.120078
##     Facebook     Whatsapp
##     3.151129     3.634497
```

4.2 Model Expansion – Interaction Effects

A positive interaction between Age and BIS15_Total (+0.1762) suggests that the protective effect of age weakens in more impulsive individuals.

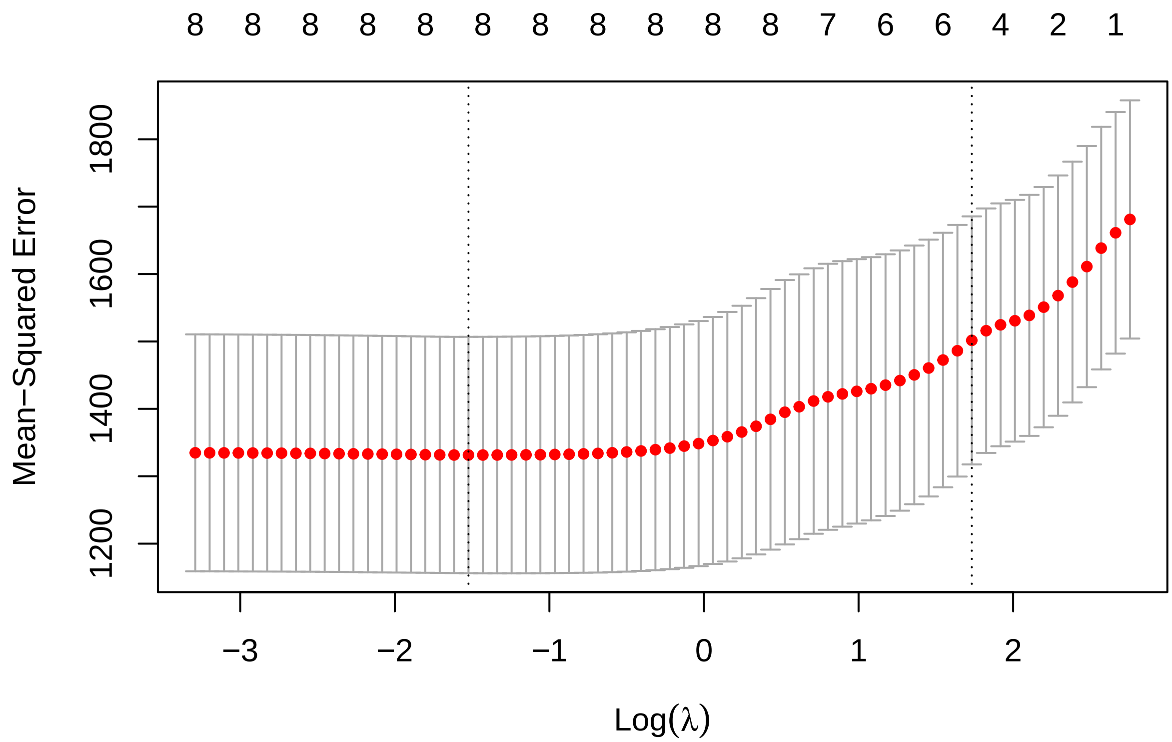
```
summary(lm(MPPUS__Total ~ BIS15_Total * Age + HADS_Total, data = df_lm))
```

```
##
## Call:
## lm(formula = MPPUS__Total ~ BIS15_Total * Age + HADS_Total, data = df_lm)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -92.111 -28.365  -2.776  28.008  74.157
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     224.8307   227.7801   0.987   0.3257
## BIS15_Total      -1.6953     7.2562  -0.234   0.8157
## Age              -9.2999    10.4762  -0.888   0.3766
## HADS_Total        1.3779     0.6358   2.167   0.0323 *
## BIS15_Total:Age   0.1762     0.3337   0.528   0.5985
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 36.74 on 114 degrees of freedom
## Multiple R-squared:  0.2234, Adjusted R-squared:  0.1962
## F-statistic:   8.2 on 4 and 114 DF,  p-value: 7.551e-06
```

4.3 Lasso Regression

The Lasso regression here is a U-shaped plot that shows the relationship between the log λ and the mean squared error (MSE). The optimal value of $\log(\lambda)$ is around -0.6, corresponding to $\lambda \approx 0.5525$, where the MSE is minimized. This implies that moderate regularization yields the best generalization performance. The two vertical dotted lines indicate the minimum MSE and the 1-SE rule (a simpler model with similar performance), and the flat MSE curve near the optimal λ implies model stability and robustness.

```
x <- model.matrix(MPPUS__Total ~ ., data = df_lm)[, -1]
y <- df_lm$MPPUS__Total
lasso_model <- cv.glmnet(x, y, alpha = 1)
plot(lasso_model)
```



4.4 Model Coefficients

Comparing the model coefficients of the linear and lasso regression model, it is found that both models consistently highlight Age, Attention, Impulsivity, Anxiety, Facebook, and WhatsApp as important, reinforcing the psychological basis of problematic smartphone use.

```r
summary(model_lm)
```

```
##
## Call:
## lm(formula = MPPUS__Total ~ CPTMeanTotal + GPA + Age + BIS15_Total +
##     HADS_Total + Instagram + Facebook + Whatsapp, data = df_lm)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -63.979 -29.407  -1.648  27.931  75.375
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  61.59518   54.14074   1.138  0.25772
## CPTMeanTotal  0.17009    0.07101   2.395  0.01829 *
## GPA           0.03553    0.02106   1.687  0.09442 .
```

```
## Age           -4.17207    1.80608  -2.310  0.02275 *
## BIS15_Total    1.33831    0.59572   2.247  0.02667 *
## HADS_Total     1.48897    0.61782   2.410  0.01761 *
## Instagram     -0.01596    0.01886  -0.846  0.39913
## Facebook       0.06441    0.02351   2.740  0.00717 **
## Whatsapp      -0.06652    0.02375  -2.801  0.00602 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34.75 on 110 degrees of freedom
## Multiple R-squared:  0.3294, Adjusted R-squared:  0.2806
## F-statistic: 6.754 on 8 and 110 DF,  p-value: 3.521e-07
```

```r
lasso_model$lambda.min
```

```
## [1] 0.2179319
```

```r
coef(lasso_model, s = "lambda.min")
```

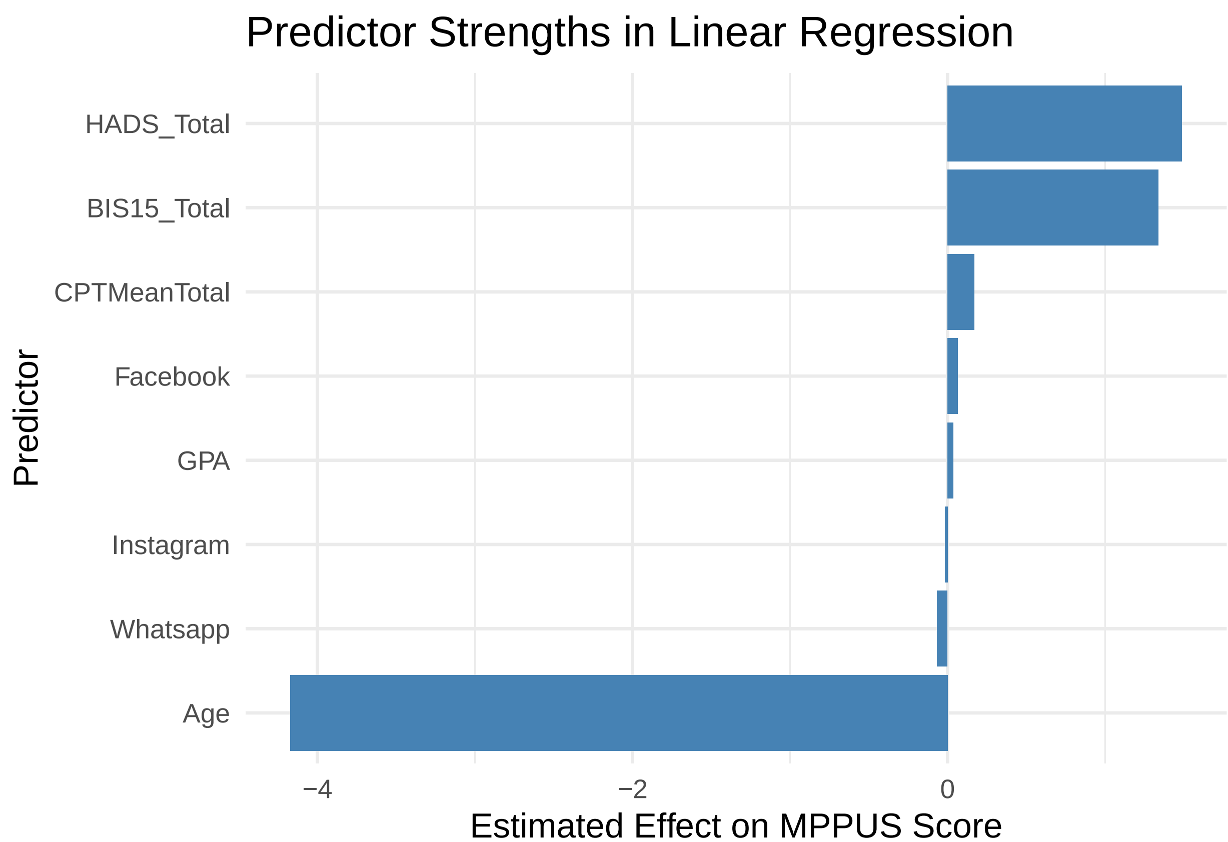
```
## 9 x 1 sparse Matrix of class "dgCMatrix"
##                       s1
## (Intercept)  61.47744414
## CPTMeanTotal  0.16329002
## GPA           0.03440366
## Age          -4.02336577
## BIS15_Total   1.34645101
## HADS_Total    1.45427823
## Instagram    -0.01436619
## Facebook      0.05869989
## Whatsapp     -0.06226279
```

4.5 Predictor Strengths

The following is a visual representation of each of the predictor's extent in predicting a problematic phone usage. It is observable that age has the strongest negative effect (-4.17), meaning younger individuals are more at risk.HADS_Total and BIS15_Total ,on the other hand, show strong positive associations (1.49 and 1.34 respectively), linking anxiety/depression and impulsivity to phone overuse. CPTMeanTotal also has a small positive effect (~0.17), supporting the attention hypothesis. Instagram and GPA, however, have negligible effects, suggesting weak predictive value.

```r
tidy(model_lm) %>%
  filter(term != "(Intercept)") %>%
  ggplot(aes(x = reorder(term, estimate), y = estimate)) +
  geom_col(fill = "steelblue") +
  coord_flip() +
```

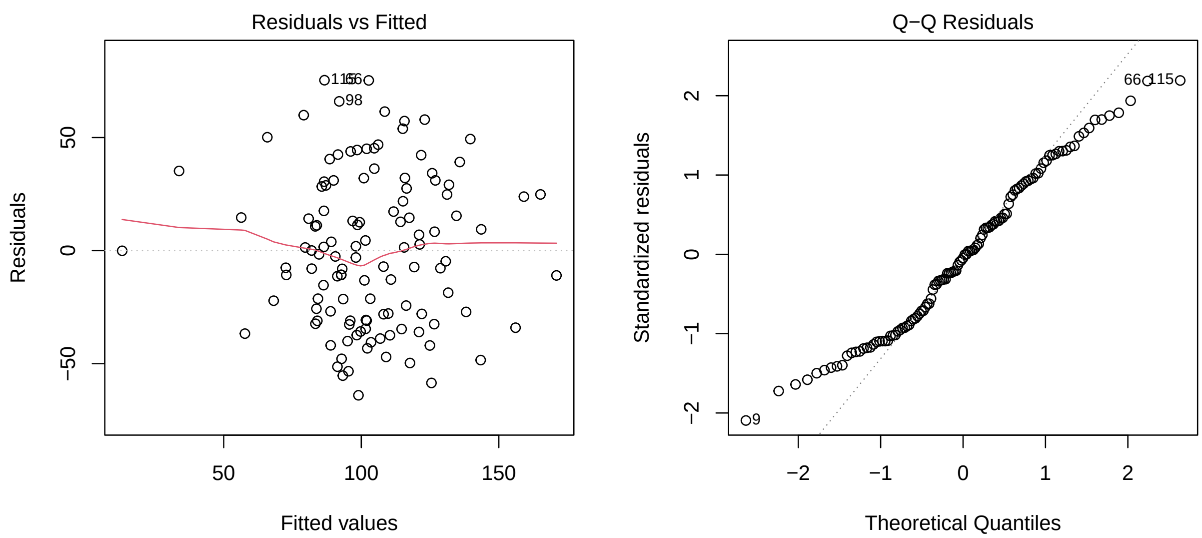
```r
  labs(
    title = "Predictor Strengths in Linear Regression",
    x = "Predictor",
    y = "Estimated Effect on MPPUS Score"
  ) +
  theme_minimal(base_size = 13)
```



4.6 Assumption Checks

It is also essential to verify the assumptions for the linear model to ensure valid and reliable results. The residual vs fitted plot shows no strong pattern, indicating linearity and constant variance are reasonably satisfied. The Q-Q plot also suggests that residuals are approximately normally distributed, supporting the appropriateness of the model.

```r
par(mfrow = c(1, 2))
plot(model_lm, which = 1)
plot(model_lm, which = 2)
```

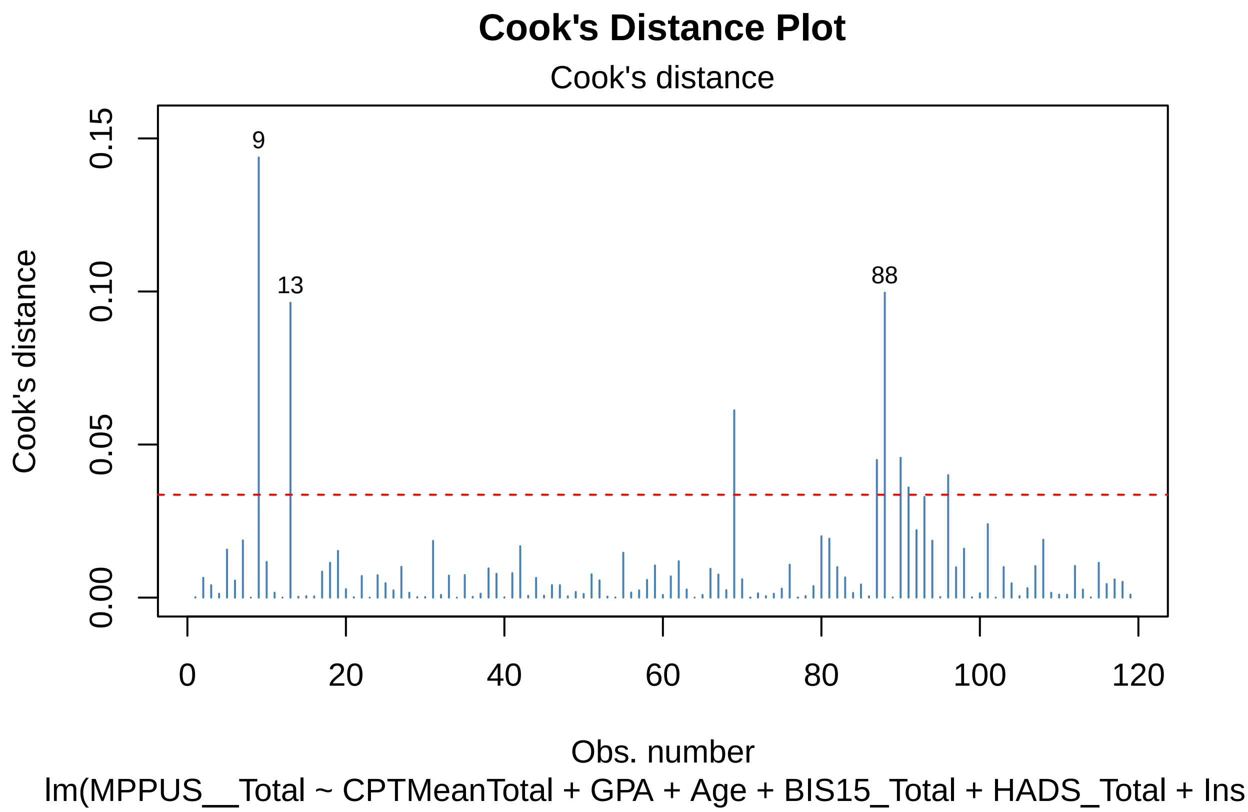



```r
par(mfrow = c(1, 1))
```

4.7 Cook's Distance Plot

Since the sample population in this study was not extremely large to be a complete representative (e.g. representing all adults), it is crucial to check if any single observation had an outsized influence on our regression results. Cooke's Distance Plot is used here and although points 9, 13, and 88 had relatively higher influence, none crossed the typical threshold — confirming the model is robust to outliers.

```r
plot(model_lm, which = 4,
     main = "Cook's Distance Plot",
     col = "steelblue",
     pch = 19,
     cex = 1)
abline(h = 4 / nrow(df_lm), col = "red", lty = 2)
```



5 Conclusion

Our analysis revealed that several psychological and behavioural traits are significant predictors of problematic smartphone use. Specifically, impulsiveness (BIS15_Total), anxiety/depression symptoms (HADS_Total), and slower attentional processing (CPTMeanTotal) showed strong positive associations with higher MPPUS scores, while age had a notable negative effect, suggesting younger individuals are more vulnerable.

Interestingly, some commonly assumed predictors like GPA and Instagram usage had negligible influence on phone addiction levels, highlighting that academic performance and platform preference may not be as strongly linked as often presumed.

These findings suggest that internal psychological traits may be more critical than external behaviours in understanding phone dependency. As smartphones become ever more integrated into our daily lives, perhaps the more important question is not *how much* we use them—but *why* we feel the need to.

-End-

6 Reference

Pluck, G. (2020). Dataset for: Cognitive Ability, Reward Processing and Personality Associated with Different Aspects of Smartphone Use. Psycharchives.org. [online] doi:https://hdl.handle.net/20.500.12034/2404.