Deciphering Career Trajectories: Insights from Aptitude and Personality Assessments

Pragya Gyawali

Overview:

In this study, we employed multiple linear regression (MLR) to analyze the relationship between aptitude, personality traits, and career outcomes. Traditional career guidance often lacks precision, so we aimed to revolutionize it by leveraging MLR alongside assessments like the OCEAN model and tests for numerical, spatial, perceptual, abstract, and verbal reasoning. Our analysis identified spatial, abstract, and verbal reasoning as key predictors of career success. By validating our findings through model selection and diagnostic checks, we offer actionable insights for individuals, career advisors, and policymakers. Our MLR approach enhances understanding and facilitates more effective career planning strategies.

Problem and Motivation:

Navigating the intricacies of career decisions presents a significant challenge for individuals seeking to align their professional paths with their inherent strengths and aspirations. Traditional approaches to career guidance often lack the precision and objectivity needed to provide tailored recommendations tailored to an individual's unique profile. In today's data-driven era, there is a growing recognition of the potential of aptitude and personality assessments to revolutionize career planning. By harnessing the power of data analytics, I aim to uncover the diverse relationships between individual attributes and career outcomes. The motivation stems from the desire to offer actionable insights that empower individuals, career advisors, and policymakers to make informed decisions and optimize career trajectories in an ever-evolving job market landscape.

Data Description:

The dataset utilized in this project originates from Kaggle and focuses on career prediction based on various aptitude and personality tests.

Variables:

OCEAN Test: The Ocean Model of Personality, commonly referred to as the Big Five personality traits, is a widely used framework in psychology to describe human personality. It assesses personality across five dimensions.

- Openness
- Conscientiousness
- Extraversion
- Agreeableness
- Neuroticism

Numerical aptitude: Evaluates an individual's proficiency in understanding and working with numbers

Spatial aptitude: The ability to mentally manipulate shapes and understand spatial relationships

Perceptual aptitude: Skills such as pattern recognition and spatial reasoning

Abstract reasoning: The understanding and manipulation of complex ideas

Verbal reasoning: Ability to comprehend and analyze written information

Research questions:

- What are the key predictors of career success, as indicated by O_score?
- How do the predictive abilities of spatial aptitude, abstract reasoning, and verbal reasoning compare in determining career success, as measured by O_score, in the final regression model?

Data Investigation

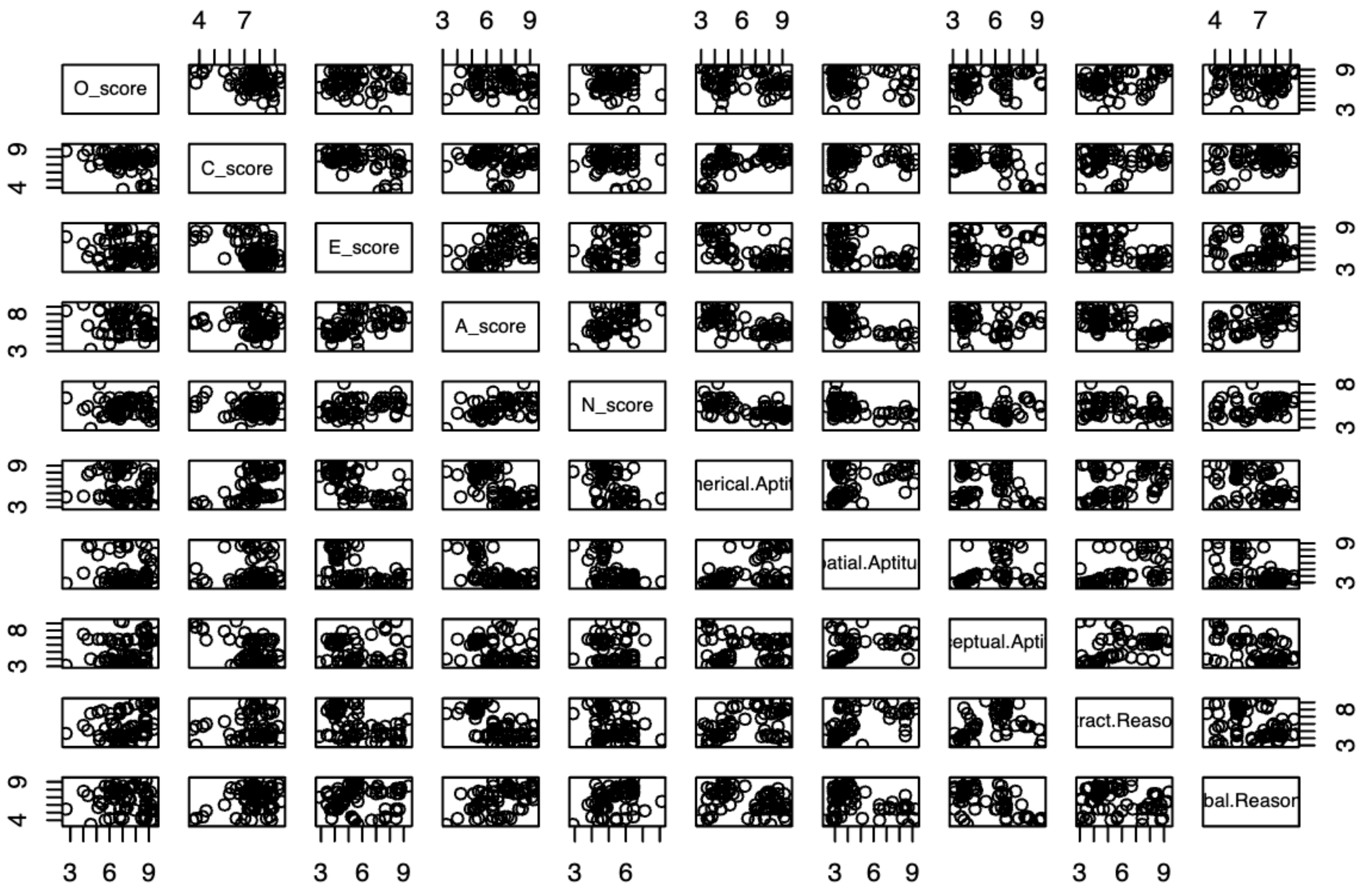
Summary of the data:

```
summary(career_data[numeric_vars])

##     O_score          C_score          E_score          A_score     
##  Min.   :2.670   Min.   :3.560   Min.   :2.890   Min.   :3.230  
##  1st Qu.:6.670   1st Qu.:7.340   1st Qu.:4.230   1st Qu.:5.450  
##  Median :7.230   Median :7.670   Median :5.230   Median :6.450  
##  Mean   :7.295   Mean   :7.538   Mean   :5.549   Mean   :6.864  
##  3rd Qu.:8.670   3rd Qu.:8.340   3rd Qu.:7.010   3rd Qu.:8.120  
##  Max.   :9.450   Max.   :9.450   Max.   :9.340   Max.   :9.340  
##     N_score       Numerical.Aptitude Spatial.Aptitude Perceptual.Aptitude
##  Min.   :2.890   Min.   :2.89       Min.   :2.340    Min.   :3.010      
##  1st Qu.:4.670   1st Qu.:4.45       1st Qu.:3.120    1st Qu.:3.670      
##  Median :5.450   Median :5.12       Median :3.450    Median :4.450      
##  Mean   :5.466   Mean   :5.94       Mean   :4.376    Mean   :5.164      
##  3rd Qu.:6.010   3rd Qu.:7.78       3rd Qu.:4.450    3rd Qu.:6.780      
##  Max.   :8.120   Max.   :9.45       Max.   :9.230    Max.   :9.340      
##  Abstract.Reasoning Verbal.Reasoning
##  Min.   :3.010      Min.   :3.450   
##  1st Qu.:4.340      1st Qu.:5.450   
##  Median :4.670      Median :7.450   
##  Mean   :5.724      Mean   :6.794   
##  3rd Qu.:7.670      3rd Qu.:8.120   
##  Max.   :9.340      Max.   :9.340   
```

Pairwise Scatter Plot

```
pairs(career_data[numeric_vars])
```



The original model:

```
summary(full_model )

##
## Call:
## lm(formula = O_score ~ ., data = career_data[, numeric_vars])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.05692 -0.59363  0.04737  0.72315  2.00666
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         15.39218    2.36032   6.521 3.36e-09 ***
## C_score             -0.62999    0.14677  -4.292 4.26e-05 ***
## E_score             -0.29727    0.08845  -3.361  0.00112 **
## A_score             -0.33289    0.11459  -2.905  0.00457 **
## N_score             -0.11198    0.16510  -0.678  0.49926
## Numerical.Aptitude  -0.18468    0.10561  -1.749  0.08358 .
## Spatial.Aptitude    -0.23037    0.07734  -2.979  0.00368 **
## Perceptual.Aptitude -0.03184    0.10982  -0.290  0.77254
## Abstract.Reasoning   0.31786    0.09426   3.372  0.00108 **
## Verbal.Reasoning     0.24259    0.10451   2.321  0.02241 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.033 on 95 degrees of freedom
## Multiple R-squared:  0.5015, Adjusted R-squared:  0.4543
## F-statistic: 10.62 on 9 and 95 DF,  p-value: 3.122e-11
```

Stepwise Model Selection:

Forward and Backwaard BIC:

```
summary(backward_bic)

##
## Call:
## lm(formula = O_score ~ C_score + E_score + A_score + Spatial.Aptitude +
##     Abstract.Reasoning + Verbal.Reasoning, data = career_data[,
##     numeric_vars])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.71843 -0.64996  0.05296  0.82018  2.09796
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        13.09627    1.50291   8.714 7.43e-14 ***
## C_score            -0.71430    0.10804  -6.611 2.01e-09 ***
## E_score            -0.23516    0.08096  -2.904 0.004547 **
## A_score            -0.27114    0.10647  -2.547 0.012435 *
## Spatial.Aptitude   -0.25044    0.07278  -3.441 0.000853 ***
## Abstract.Reasoning  0.31983    0.08057   3.969 0.000137 ***
## Verbal.Reasoning    0.29647    0.09424   3.146 0.002194 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.034 on 98 degrees of freedom
## Multiple R-squared:  0.4853, Adjusted R-squared:  0.4538
## F-statistic:  15.4 on 6 and 98 DF,  p-value: 2.3e-12
```

```
summary(forward_bic)

##
## Call:
## lm(formula = O_score ~ Abstract.Reasoning + C_score + Spatial.Aptitude +
##     E_score + Verbal.Reasoning + A_score, data = career_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.71843 -0.64996  0.05296  0.82018  2.09796
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        13.09627    1.50291   8.714 7.43e-14 ***
## Abstract.Reasoning  0.31983    0.08057   3.969 0.000137 ***
## C_score            -0.71430    0.10804  -6.611 2.01e-09 ***
## Spatial.Aptitude   -0.25044    0.07278  -3.441 0.000853 ***
## E_score            -0.23516    0.08096  -2.904 0.004547 **
## Verbal.Reasoning    0.29647    0.09424   3.146 0.002194 **
## A_score            -0.27114    0.10647  -2.547 0.012435 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.034 on 98 degrees of freedom
## Multiple R-squared:  0.4853, Adjusted R-squared:  0.4538
## F-statistic:  15.4 on 6 and 98 DF,  p-value: 2.3e-12
```

Forward and Backward AIC

```r
summary(backward_aic)
```

```
##
## Call:
## lm(formula = O_score ~ C_score + E_score + A_score + Numerical.Aptitude +
##     Spatial.Aptitude + Abstract.Reasoning + Verbal.Reasoning,
##     data = career_data[, numeric_vars])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.79590 -0.55346  0.09282  0.72294  1.92204
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        14.29370    1.66408   8.590 1.48e-13 ***
## C_score            -0.61058    0.12485  -4.890 3.99e-06 ***
## E_score            -0.28696    0.08644  -3.320  0.00127 **
## A_score            -0.33783    0.11335  -2.980  0.00364 **
## Numerical.Aptitude -0.15585    0.09627  -1.619  0.10874
## Spatial.Aptitude   -0.22229    0.07425  -2.994  0.00350 **
## Abstract.Reasoning  0.30366    0.08054   3.770  0.00028 ***
## Verbal.Reasoning    0.24657    0.09843   2.505  0.01391 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.025 on 97 degrees of freedom
## Multiple R-squared:  0.4989, Adjusted R-squared:  0.4627
## F-statistic: 13.79 on 7 and 97 DF,  p-value: 2.797e-12
```

```r
summary(forward_aic)
```

```
##
## Call:
## lm(formula = O_score ~ Abstract.Reasoning + C_score + Spatial.Aptitude +
##     E_score + Verbal.Reasoning + A_score + Numerical.Aptitude,
##     data = career_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.79590 -0.55346  0.09282  0.72294  1.92204
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        14.29370    1.66408   8.590 1.48e-13 ***
## Abstract.Reasoning  0.30366    0.08054   3.770  0.00028 ***
## C_score            -0.61058    0.12485  -4.890 3.99e-06 ***
## Spatial.Aptitude   -0.22229    0.07425  -2.994  0.00350 **
## E_score            -0.28696    0.08644  -3.320  0.00127 **
## Verbal.Reasoning    0.24657    0.09843   2.505  0.01391 *
## A_score            -0.33783    0.11335  -2.980  0.00364 **
## Numerical.Aptitude -0.15585    0.09627  -1.619  0.10874
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.025 on 97 degrees of freedom
## Multiple R-squared:  0.4989, Adjusted R-squared:  0.4627
## F-statistic: 13.79 on 7 and 97 DF,  p-value: 2.797e-12
```

Model Selection using Anova:

```
anova(backward_bic, forward_bic, backward_aic, forward_aic)
```

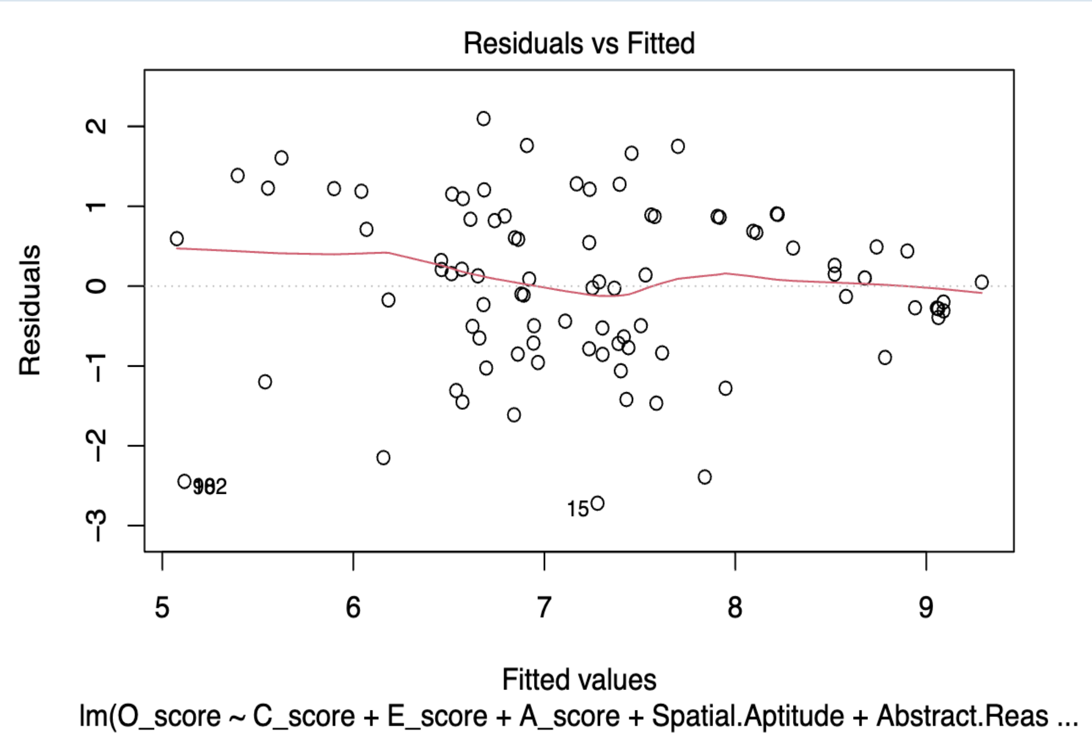
```
## Analysis of Variance Table
##
## Model 1: O_score ~ C_score + E_score + A_score + Spatial.Aptitude + Abstract.Reasoning +
##     Verbal.Reasoning
## Model 2: O_score ~ Abstract.Reasoning + C_score + Spatial.Aptitude + E_score +
##     Verbal.Reasoning + A_score
## Model 3: O_score ~ C_score + E_score + A_score + Numerical.Aptitude +
##     Spatial.Aptitude + Abstract.Reasoning + Verbal.Reasoning
## Model 4: O_score ~ Abstract.Reasoning + C_score + Spatial.Aptitude + E_score +
##     Verbal.Reasoning + A_score + Numerical.Aptitude
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     98 104.69
## 2     98 104.69  0    0.0000
## 3     97 101.93  1    2.7538 2.6205 0.1087
## 4     97 101.93  0    0.0000
```

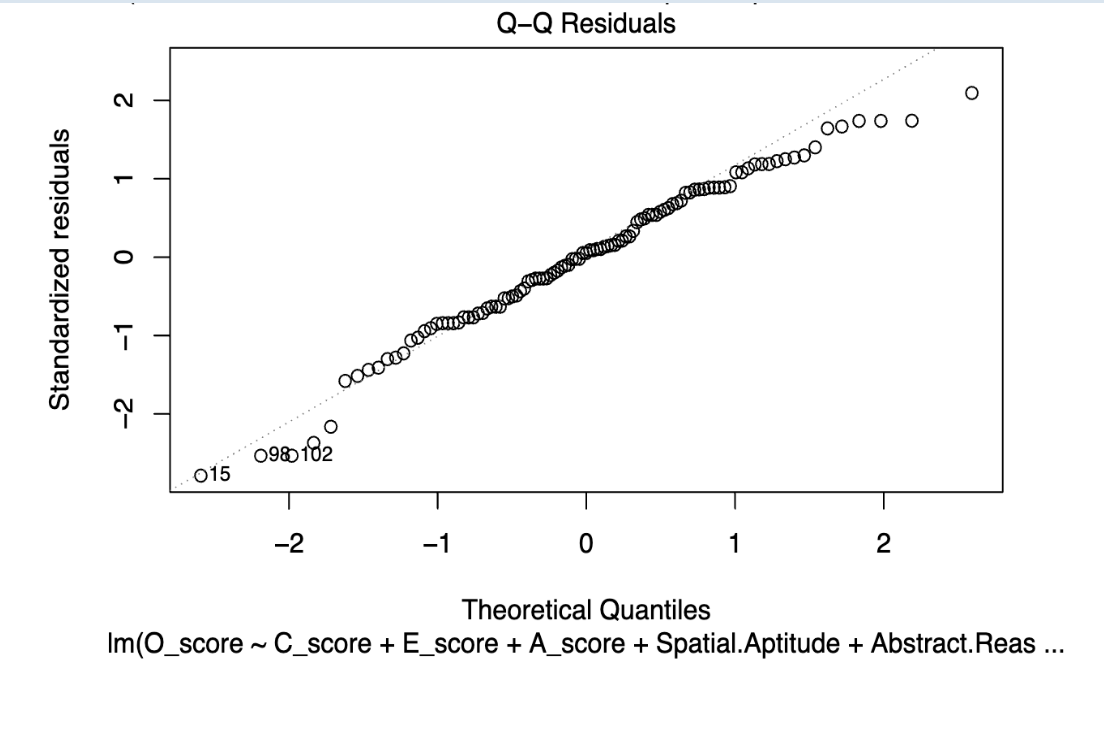
Final Model:

```r
summary(final_model)
```

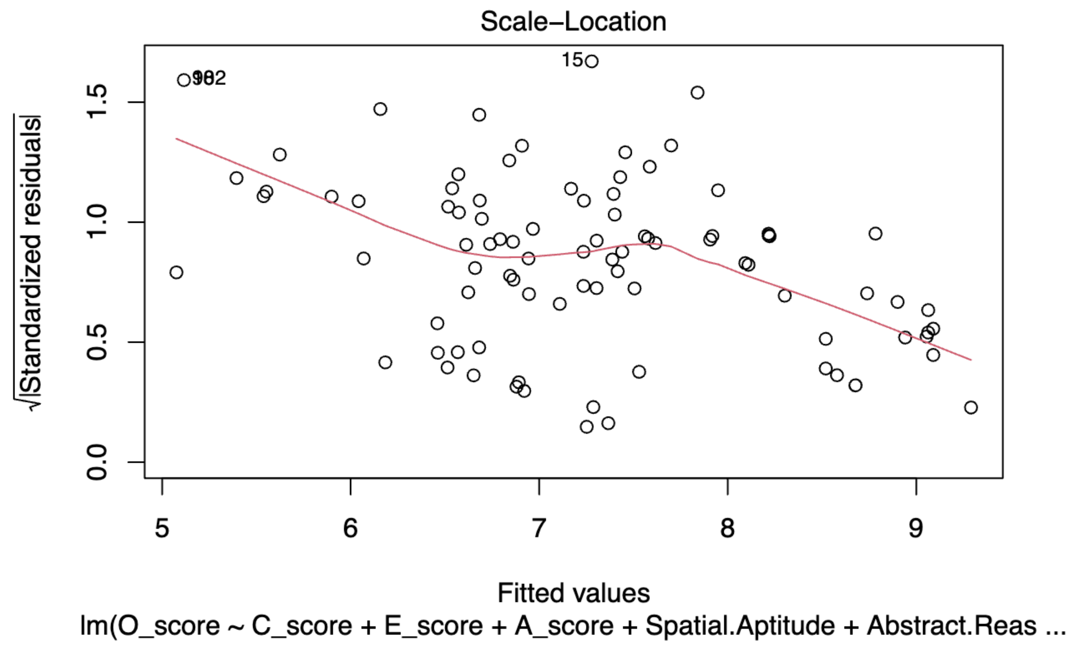
```
##
## Call:
## lm(formula = O_score ~ C_score + E_score + A_score + Spatial.Aptitude +
##     Abstract.Reasoning + Verbal.Reasoning, data = career_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.71843 -0.64996  0.05296  0.82018  2.09796
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        13.09627    1.50291   8.714 7.43e-14 ***
## C_score            -0.71430    0.10804  -6.611 2.01e-09 ***
## E_score            -0.23516    0.08096  -2.904 0.004547 **
## A_score            -0.27114    0.10647  -2.547 0.012435 *
## Spatial.Aptitude   -0.25044    0.07278  -3.441 0.000853 ***
## Abstract.Reasoning  0.31983    0.08057   3.969 0.000137 ***
## Verbal.Reasoning    0.29647    0.09424   3.146 0.002194 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.034 on 98 degrees of freedom
## Multiple R-squared:  0.4853, Adjusted R-squared:  0.4538
## F-statistic:  15.4 on 6 and 98 DF,  p-value: 2.3e-12
```

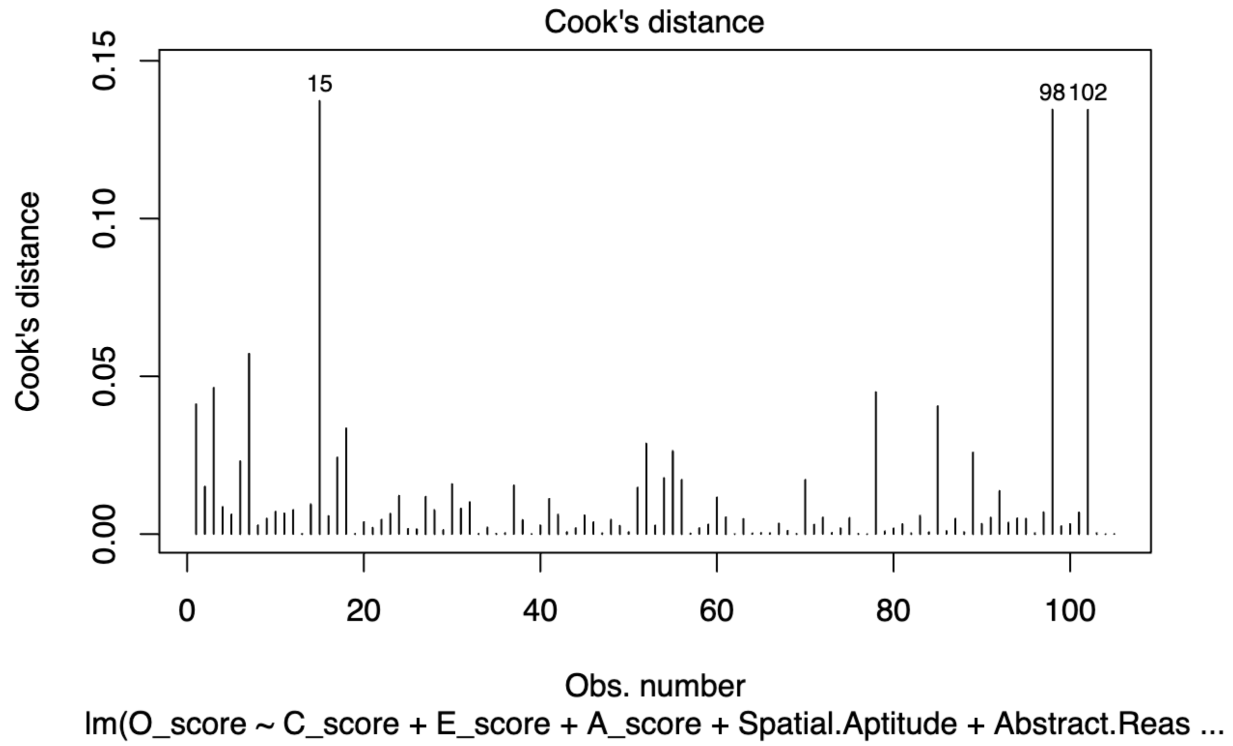
Diagnostic Check for Final Model:



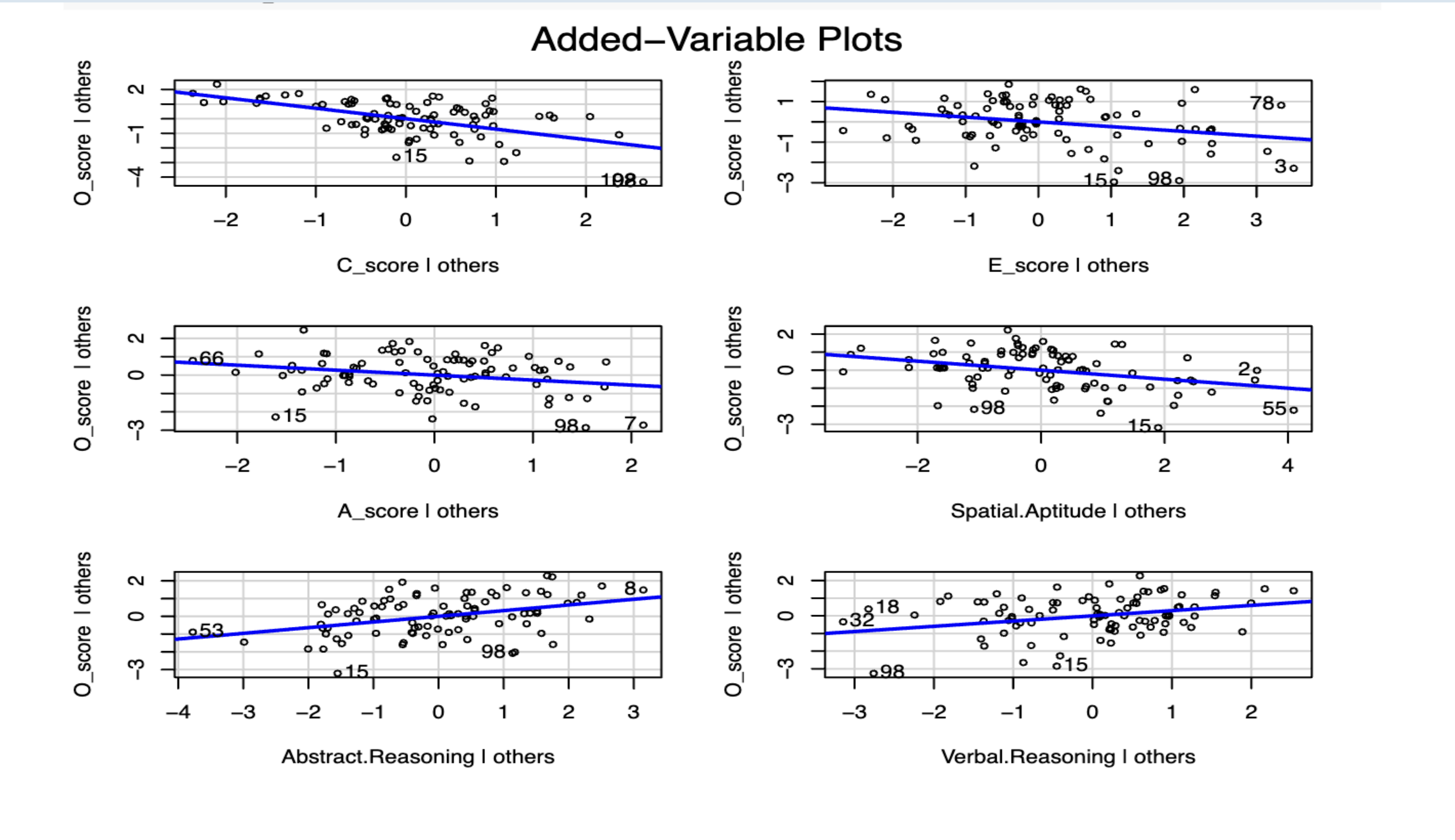



```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(final_model)
## W = 0.98169, p-value = 0.1563
```


Scale−Location
√|Standardized residuals|
Fitted values
lm(O_score ~ C_score + E_score + A_score + Spatial.Aptitude + Abstract.Reas ...


Cook's distance
Cook's distance
Obs. number
lm(O_score ~ C_score + E_score + A_score + Spatial.Aptitude + Abstract.Reas ...

Added Variable Plots:



MLR Equation

O_score:

$$13.09627-0.71430*\text{C\_score}-0.23516*\text{E\_score}-0.27114*\text{A\_score}-0.25044\times\text{Spatial.Aptitude}+0.31983*\text{Abstract.Reasoning}+0.29647*\text{Verbal.Reasoning}$$

Findings:

- The key predictors of career success, as indicated by O_score, are Spatial Aptitude, Abstract Reasoning, and Verbal Reasoning. These factors demonstrate significant associations with career success, suggesting that individuals who perform well in spatial visualization, abstract thinking, and verbal comprehension are more likely to excel in their chosen careers.
- When comparing the predictive abilities of specific aptitude and reasoning factors in determining career success, our analysis reveals that Spatial Aptitude emerges as the strongest predictor, followed closely by Abstract Reasoning and Verbal Reasoning. While all three factors exhibit significant associations with career success, Spatial Aptitude appears to have the most substantial impact, indicating that individuals with strong spatial visualization skills are particularly well-suited for successful careers.

Conclusion:

Our findings revealed significant associations between various predictor variables and career success, with Spatial Aptitude, Abstract Reasoning, and Verbal Reasoning emerging as key predictors. These results underscore the importance of considering individual strengths and attributes in career planning processes, empowering individuals to make informed decisions aligned with their capabilities and aspirations. In conclusion, our report highlights the pivotal role of multiple linear regression in elucidating the predictive power of aptitude and personality traits in career outcomes.

Thank you