02 - Regression Analysis

Alec Stashevsky

11/27/2021

Build Regression Data

```r
# Keep only columns we need and coerce as ordinal / categorical
reg.data <- post[, .(
  L3 = factor(L3, ordered = TRUE),
  L2 = factor(L2, ordered = TRUE),
  L1 = factor(L1, ordered = TRUE),
  D1 = factor(D1, ordered = FALSE),
  D4 = factor(D4, ordered = FALSE),
  D5 = factor(D5, ordered = FALSE),
  D6 = factor(D6, ordered = FALSE)
  )][,
    Delta := as.numeric(L3) - as.numeric(L2)
  ]

# Explore
summary(reg.data)
```

```
##  L3     L2     L1     D1        D4      D5     D6          Delta
##  0: 2   0: 3   0: 2   0: 5   1   :29   0:26   0:36   Min.   :-1.0000
##  1: 6   1: 7   1: 7   1:46   2   :14   1:28   1:41   1st Qu.: 0.0000
##  2: 3   2:10   2:10   2:27   3   :18   2: 9   2: 1   Median : 0.0000
##  3:12   3:17   3:25          4   : 5   3: 8          Mean   : 0.5385
##  4:30   4:32   4:34          5   :11   4: 7          3rd Qu.: 1.0000
##  5:25   5: 9                 NA's: 1                 Max.   : 2.0000
```

There is one missing value for respondent 37 on Question D4. This respondent may be dropped from the regression, lets make sure to check.

Regression Analysis - Importance of Teaching Climate Change

Linear Regression

We will first test run a multivariate linear regression where we treat the independent variable as continuous. Our first regression will be of the form

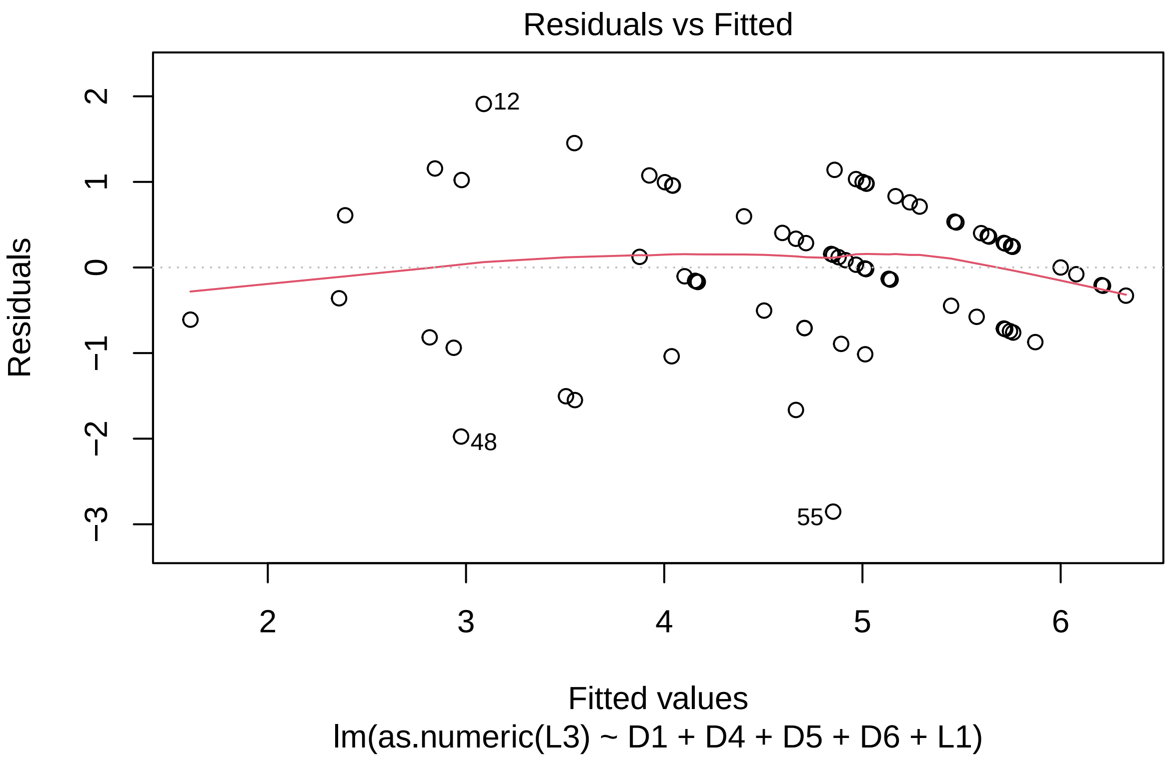
$$L3_i = D1_i + D4_i + D5_i + D6_i + L1_i + \epsilon_i$$

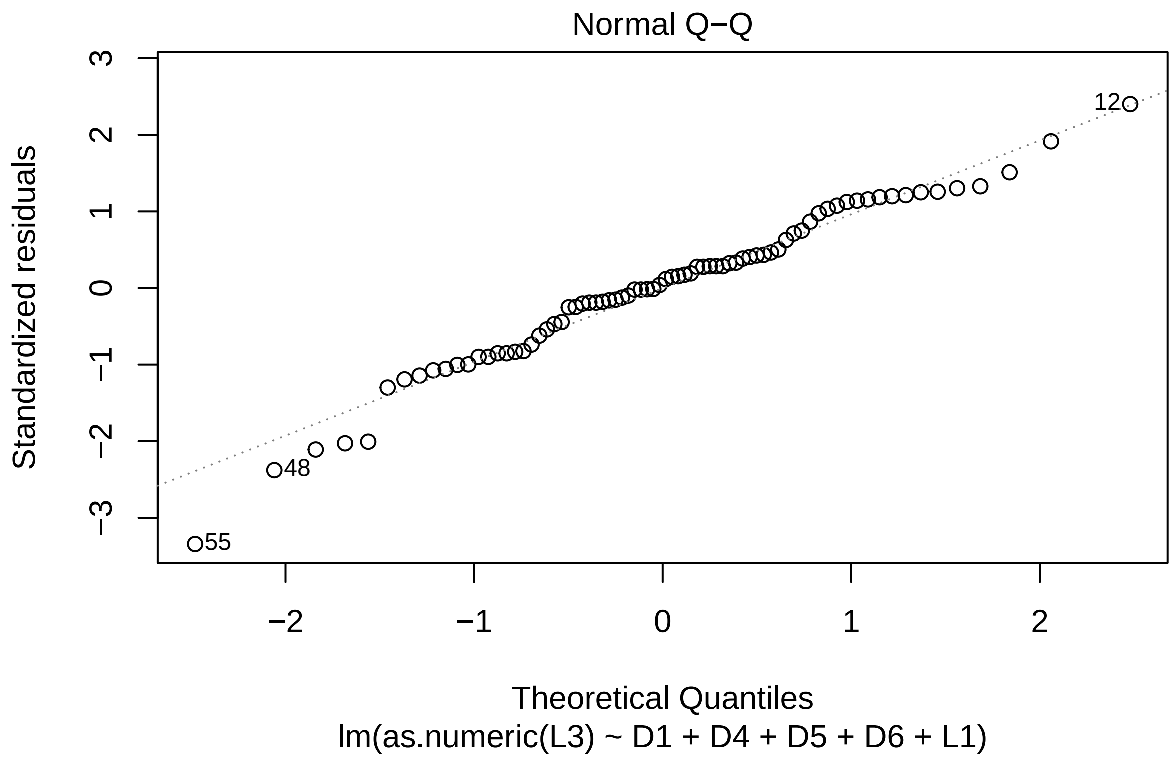
where DX indicates categorical demographic variables and LX indicates ordinal Likert scale variables.

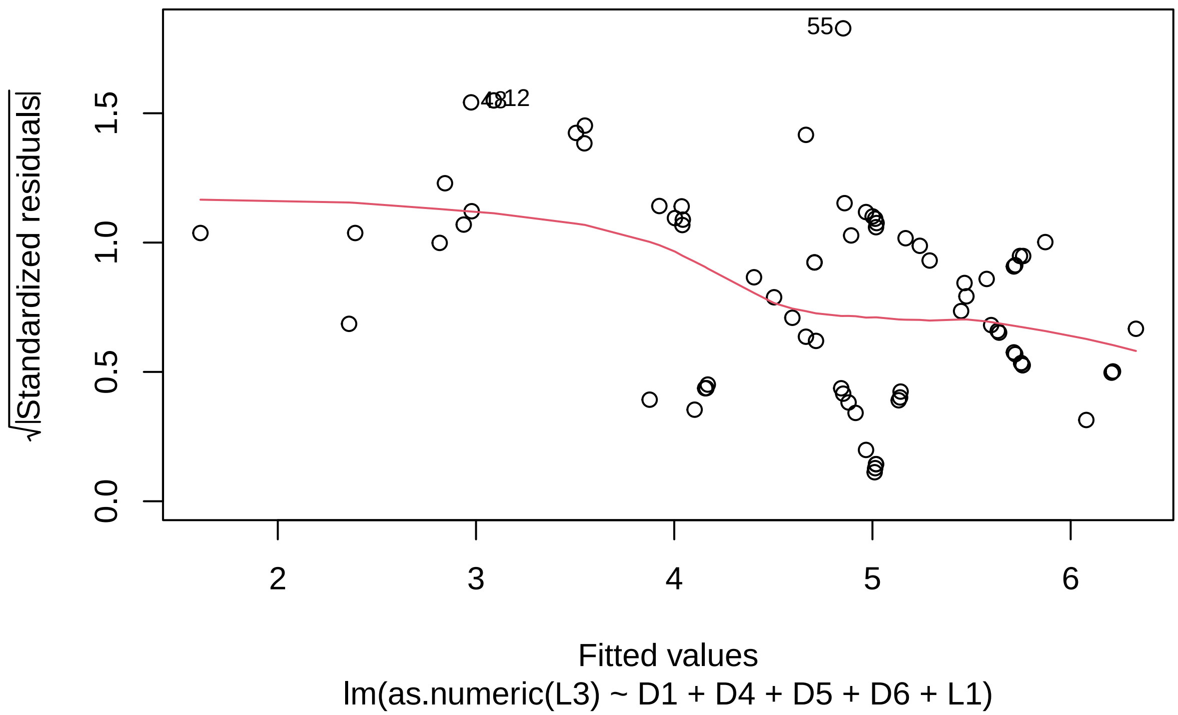
```r
# Model 1 - we treat independent variable as continuous
mlr1 <- lm(as.numeric(L3) ~ D1 + D4 + D5 + D6 + L1, data = reg.data)
mlr1.table <- broom::tidy(mlr1)
summary(mlr1)
```

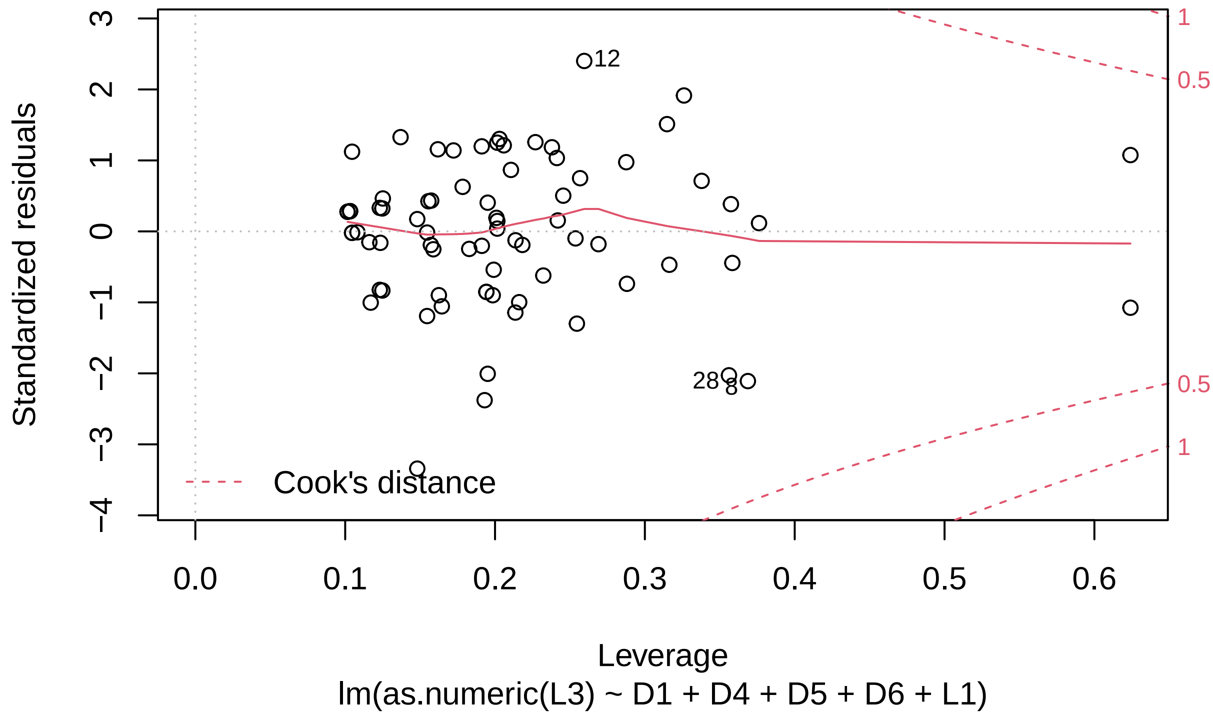
```
##
## Call:
## lm(formula = as.numeric(L3) ~ D1 + D4 + D5 + D6 + L1, data = reg.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.85233 -0.50379  0.03249  0.52603  1.91045
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.761439   0.616920   6.097 8.45e-08 ***
## D11          0.295777   0.547764   0.540    0.591
## D12         -0.128991   0.481548  -0.268    0.790
## D42         -0.118515   0.331210  -0.358    0.722
## D43         -0.158640   0.290014  -0.547    0.586
## D44         -0.615535   0.468412  -1.314    0.194
## D45          0.334547   0.358522   0.933    0.354
## D51         -0.007011   0.293092  -0.024    0.981
## D52          0.411491   0.572116   0.719    0.475
## D53         -0.453252   0.556065  -0.815    0.418
## D54         -0.549863   0.606765  -0.906    0.368
## D61          0.121441   0.229831   0.528    0.599
## D62          0.242449   0.970472   0.250    0.804
## L1.L         2.714093   0.482005   5.631 5.03e-07 ***
## L1.Q         0.107345   0.461424   0.233    0.817
## L1.C        -0.257982   0.359610  -0.717    0.476
## L1^4         0.066886   0.314968   0.212    0.833
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9246 on 60 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.6063, Adjusted R-squared:  0.5014
## F-statistic: 5.776 on 16 and 60 DF,  p-value: 2.582e-07
```

```r
plot(mlr1, caption = list("Residuals vs Fitted", "Normal Q-Q"))
```


Residuals vs Fitted
Residuals
2
1
0
−1
−2
−3
12
48
55
2
3
4
5
6
Fitted values
lm(as.numeric(L3) ~ D1 + D4 + D5 + D6 + L1)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(as.numeric(L3) ~ D1 + D4 + D5 + D6 + L1)
12
48
55
−3
−2
−1
0
1
2
3
−2
−1
0
1
2


55
48
12
0.0
0.5
1.0
1.5
√|Standardized residuals|
2
3
4
5
6
Fitted values
lm(as.numeric(L3) ~ D1 + D4 + D5 + D6 + L1)




```r
# Run ANOVA
mlr1.anova <- anova(mlr1)
mlr1.anova
```

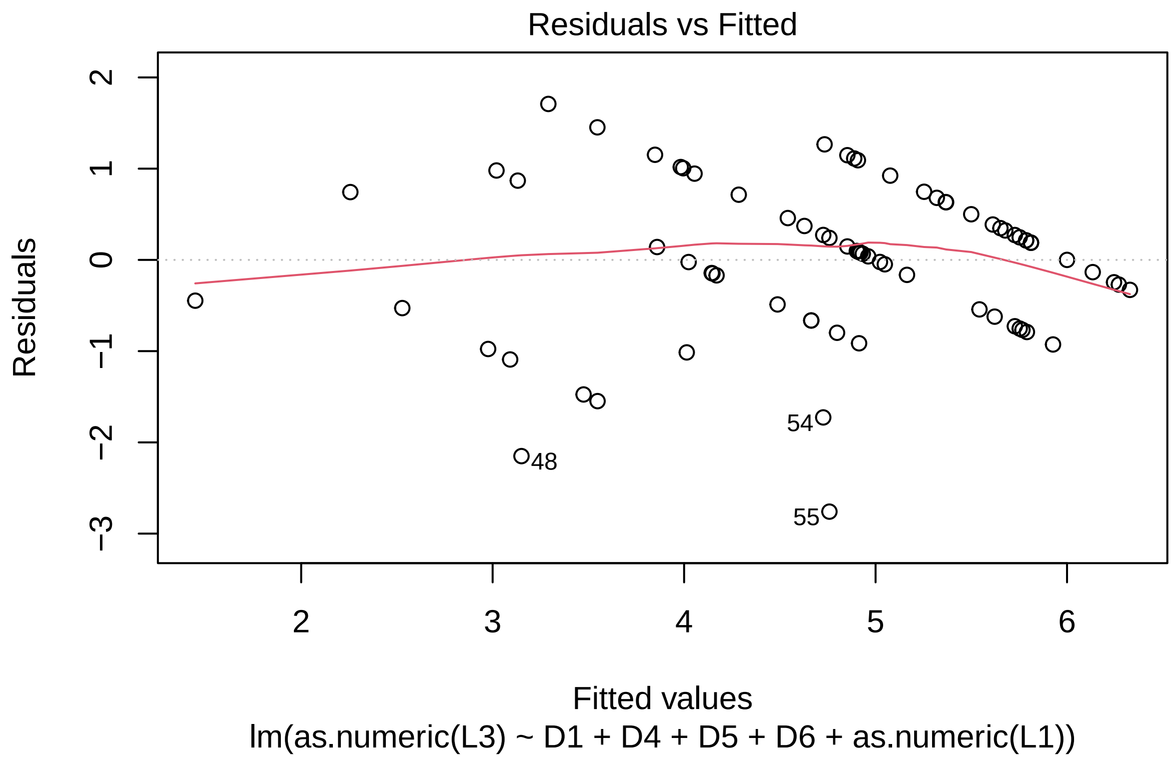
```
## Analysis of Variance Table
##
## Response: as.numeric(L3)
##           Df Sum Sq Mean Sq F value    Pr(>F)
## D1         2  6.149  3.0744  3.5959    0.0335 *
## D4         4  6.678  1.6694  1.9526    0.1134
## D5         4  2.176  0.5440  0.6362    0.6386
## D6         2  1.888  0.9441  1.1042    0.3381
## L1         4 62.124 15.5309 18.1657 8.006e-10 ***
## Residuals 60 51.298  0.8550
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

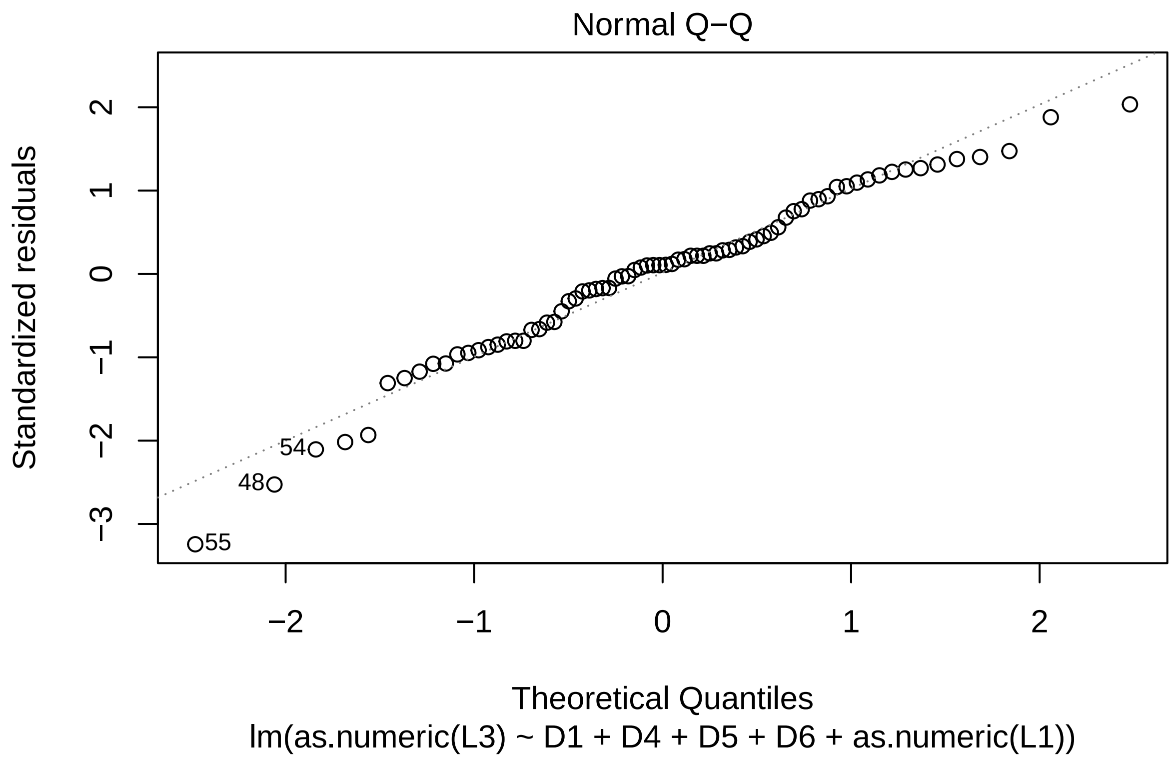
```r
# Model 2 - we treat all Likert variables as continuous
mlr2 <- lm(as.numeric(L3) ~ D1 + D4 + D5 + D6 + as.numeric(L1), data = reg.data)
mlr2.table <- broom::tidy(mlr2)
summary(mlr2)
```

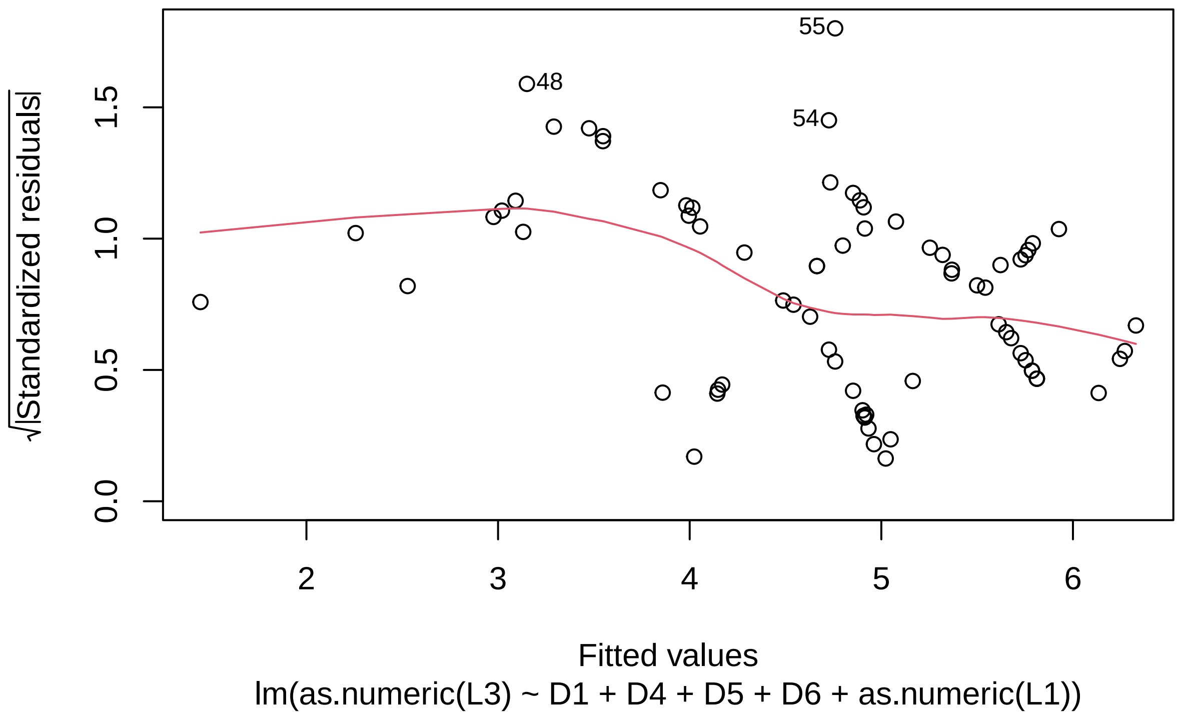
```
##
## Call:
## lm(formula = as.numeric(L3) ~ D1 + D4 + D5 + D6 + as.numeric(L1),
```

```
##     data = reg.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.75906 -0.52822  0.09215  0.50031  1.70900
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.08858    0.73086   1.489    0.141
## D11              0.30480    0.52456   0.581    0.563
## D12             -0.07943    0.46789  -0.170    0.866
## D42             -0.13434    0.32247  -0.417    0.678
## D43             -0.17414    0.28300  -0.615    0.541
## D44             -0.62240    0.45887  -1.356    0.180
## D45              0.34384    0.34386   1.000    0.321
## D51              0.02535    0.27989   0.091    0.928
## D52              0.38816    0.53138   0.730    0.468
## D53             -0.44085    0.53878  -0.818    0.416
## D54             -0.54121    0.58624  -0.923    0.359
## D61              0.11504    0.21756   0.529    0.599
## D62              0.21354    0.94854   0.225    0.823
## as.numeric(L1)   0.87862    0.10161   8.647 2.66e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9074 on 63 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.602,  Adjusted R-squared:  0.5198
## F-statistic: 7.329 on 13 and 63 DF,  p-value: 1.774e-08
```

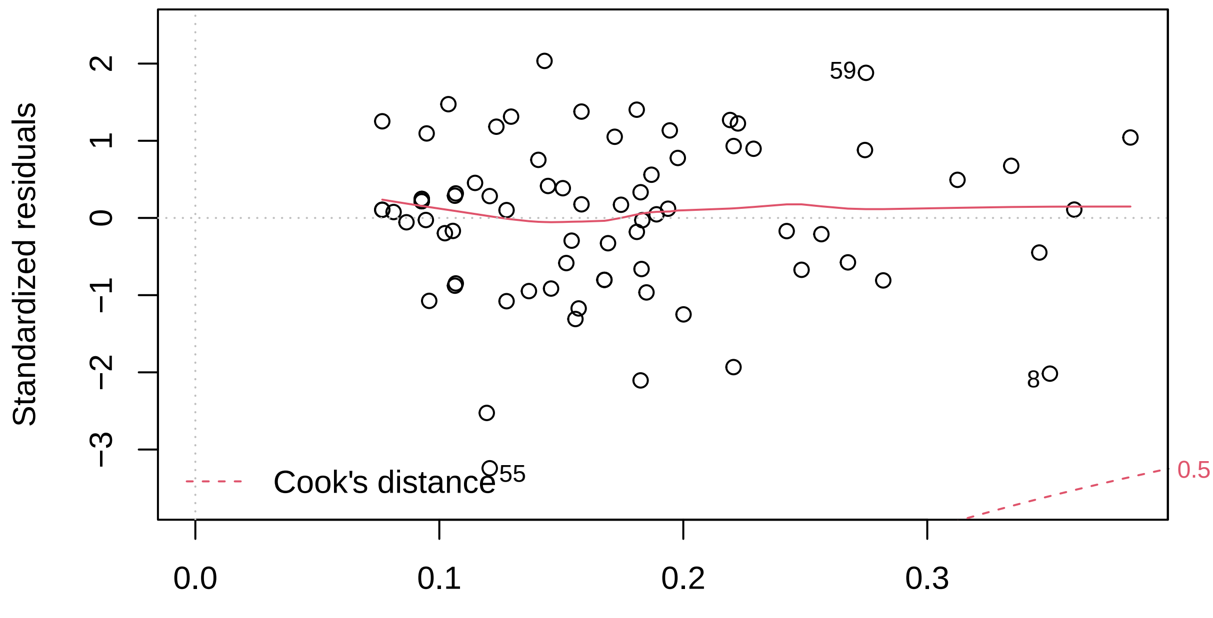
```r
plot(mlr2, caption = list("Residuals vs Fitted", "Normal Q-Q"))
```


Residuals vs Fitted
Residuals
Fitted values
lm(as.numeric(L3) ~ D1 + D4 + D5 + D6 + as.numeric(L1))
48
54
55


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(as.numeric(L3) ~ D1 + D4 + D5 + D6 + as.numeric(L1))
54
48
55
−3
−2
−1
0
1
2



55
48
54
1.5
1.0
0.5
0.0
√|Standardized residuals|
2
3
4
5
6
Fitted values
lm(as.numeric(L3) ~ D1 + D4 + D5 + D6 + as.numeric(L1))



lm(as.numeric(L3) ~ D1 + D4 + D5 + D6 + as.numeric(L1))

```r
# Run ANOVA
mlr2.anova <- anova(mlr2)
mlr2.anova
```

```
## Analysis of Variance Table
##
## Response: as.numeric(L3)
##                  Df Sum Sq Mean Sq F value    Pr(>F)
## D1                2  6.149   3.074  3.7342   0.02934 *
## D4                4  6.678   1.669  2.0277   0.10124
## D5                4  2.176   0.544  0.6607   0.62160
## D6                2  1.888   0.944  1.1467   0.32423
## as.numeric(L1)    1 61.553  61.553 74.7626 2.657e-12 ***
## Residuals        63 51.869   0.823
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ordinal Logistic Regression

Now, we will proceed with an ordinal logistic regression of the form

$$logit(P(Y \leq j)) = \beta_{j0} - \eta_1 x_1 - \cdots - \eta_p x_p$$

where Y is an ordinal Likert variable with J categories. $P(Y \leq j)$ represents the cumulative probability of Y being less than or equal to a specific category $j = 1, \ldots, J-1$. In our case $J = 5$ and the response and predictor variables are the same as our linear regression specification.

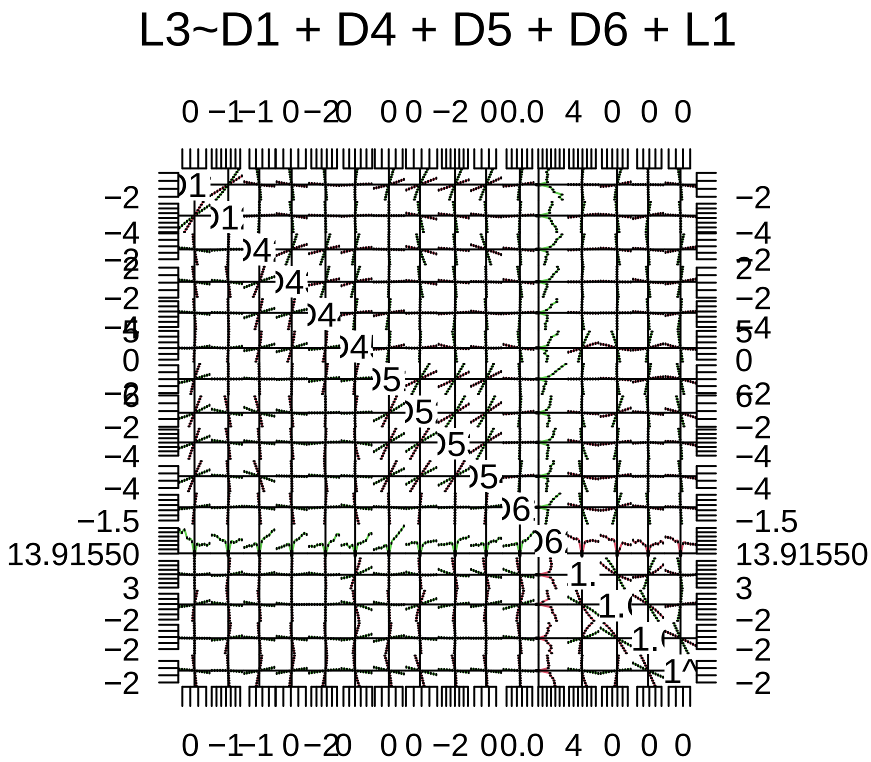
```r
# Model 1 - we treat independent variable as continuous
olr1 <- polr(L3 ~ D1 + D4 + D5 + D6 + L1, data = reg.data, Hess = TRUE)
olr1.table <- broom::tidy(olr1)

# Test the assumptions for proportional-odds
brant(olr1)
```

```
## --------------------------------------------
## Test for X2  df  probability
## --------------------------------------------
## Omnibus      -96.59  64  1
## D11      0   4   1
## D12      0   4   1
## D42      0   4   1
## D43      0   4   1
## D44      0   4   1
## D45      4.09    4   0.39
## D51      0   4   1
## D52      0   4   1
## D53      0   4   1
## D54      0   4   1
## D61      -0.11   4   1
## D62      0   4   1
## L1.L     0   4   1
## L1.Q     0   4   1
## L1.C     0   4   1
## L1^4     0   4   1
## --------------------------------------------
##
## H0: Parallel Regression Assumption holds
```

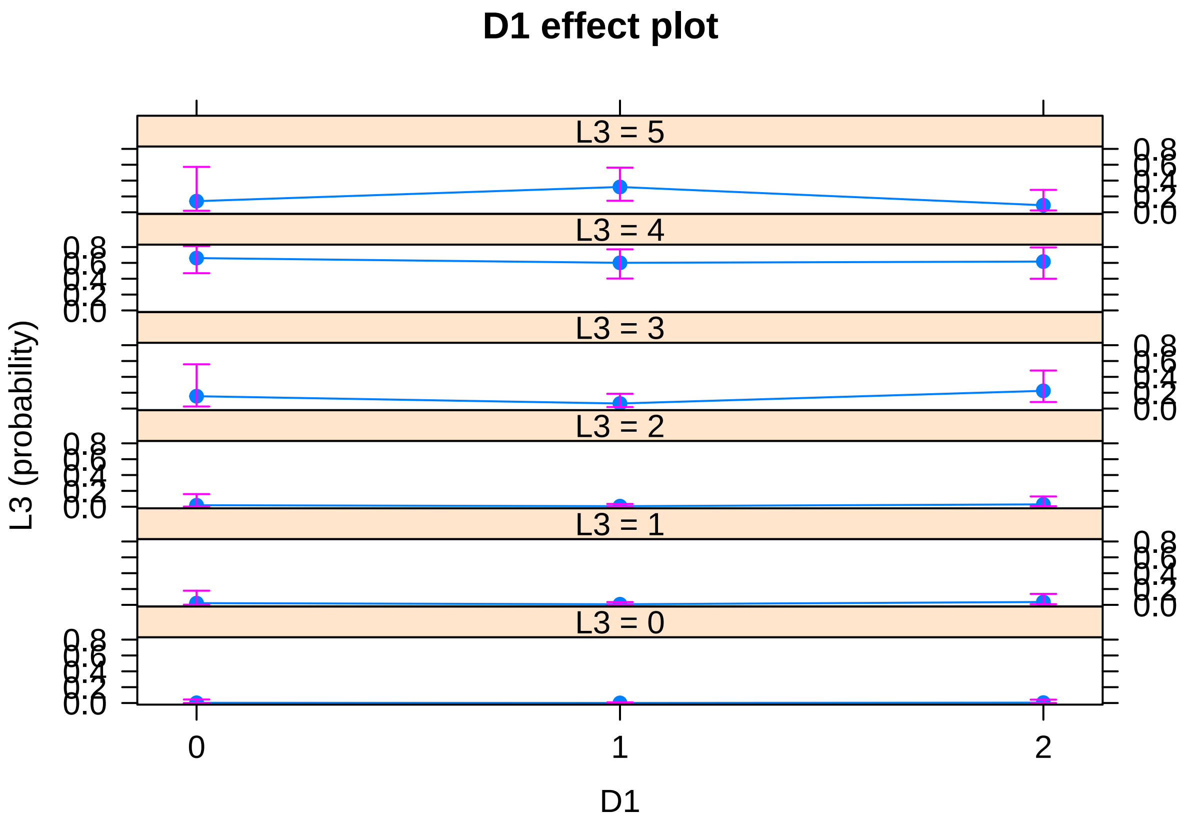
```r
# Check goodness of fit
paste("Goodness of fit Chi-sq:", 1-pchisq(deviance(olr1),df.residual(olr1)))
```

```
## [1] "Goodness of fit Chi-sq: 1.58394630744851e-10"
```

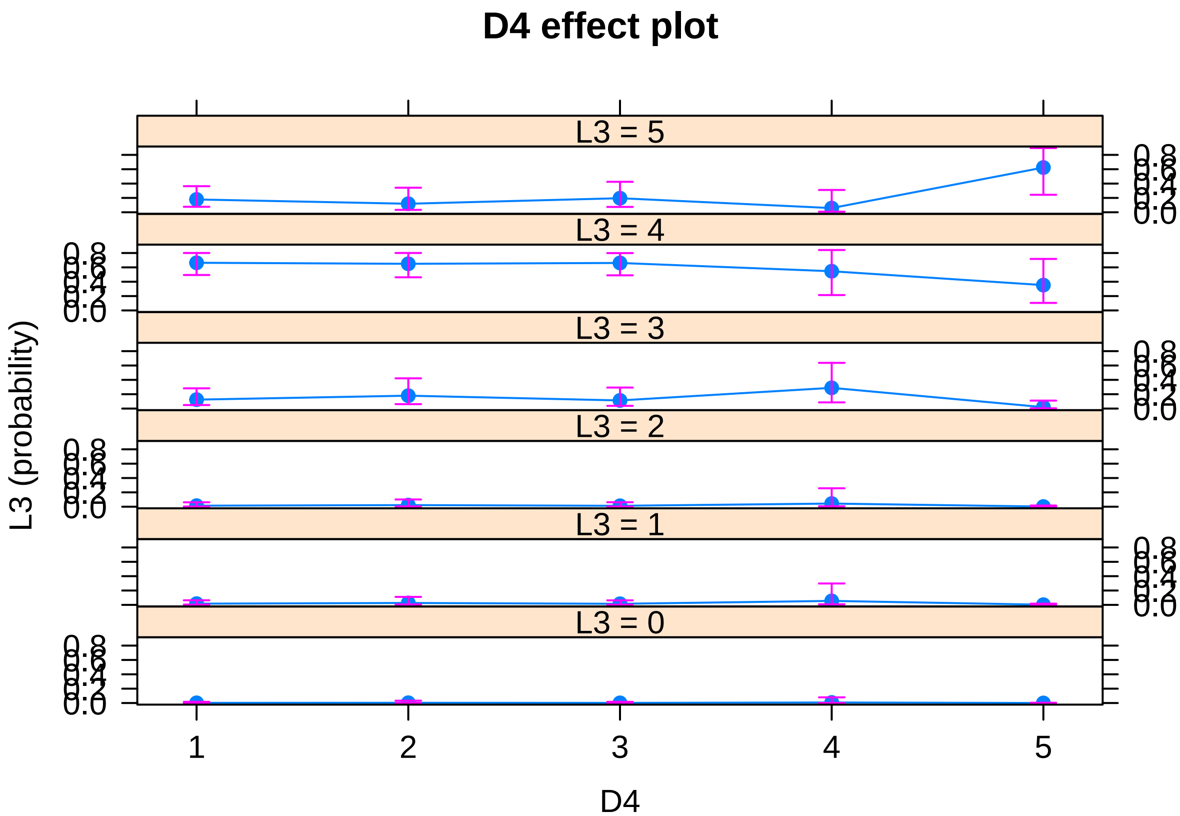
```r
# ANOVA
olr1.anova <- broom::tidy(Anova(olr1))

# Add p-values to regression table
olr1.table$p.value <- pnorm(abs(olr1.table$statistic), lower.tail = FALSE) * 2

# Diagnostic plots
olr1.pr <- profile(olr1)
pairs(olr1.pr)
```



```r
predictors <- c("D1", "D4", "D5", "D6", "L1")
for (p in predictors) {
  print(plot(effects::Effect(focal.predictors = c(p), mod = olr1)))
}
```

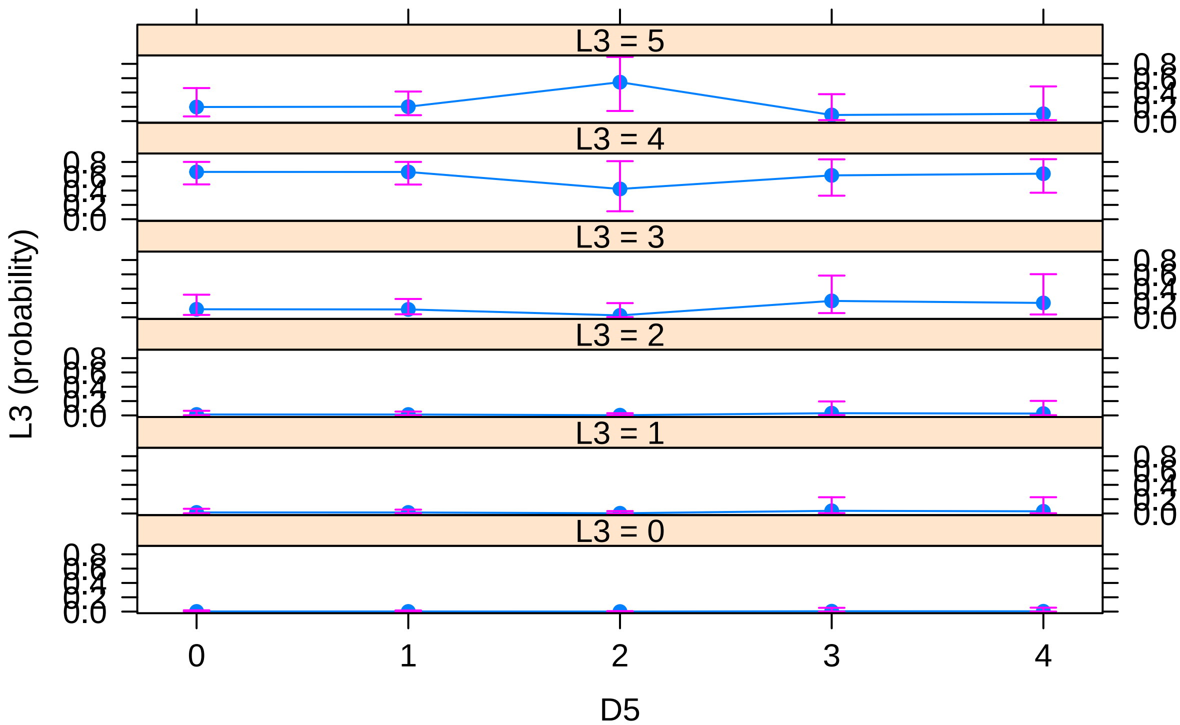


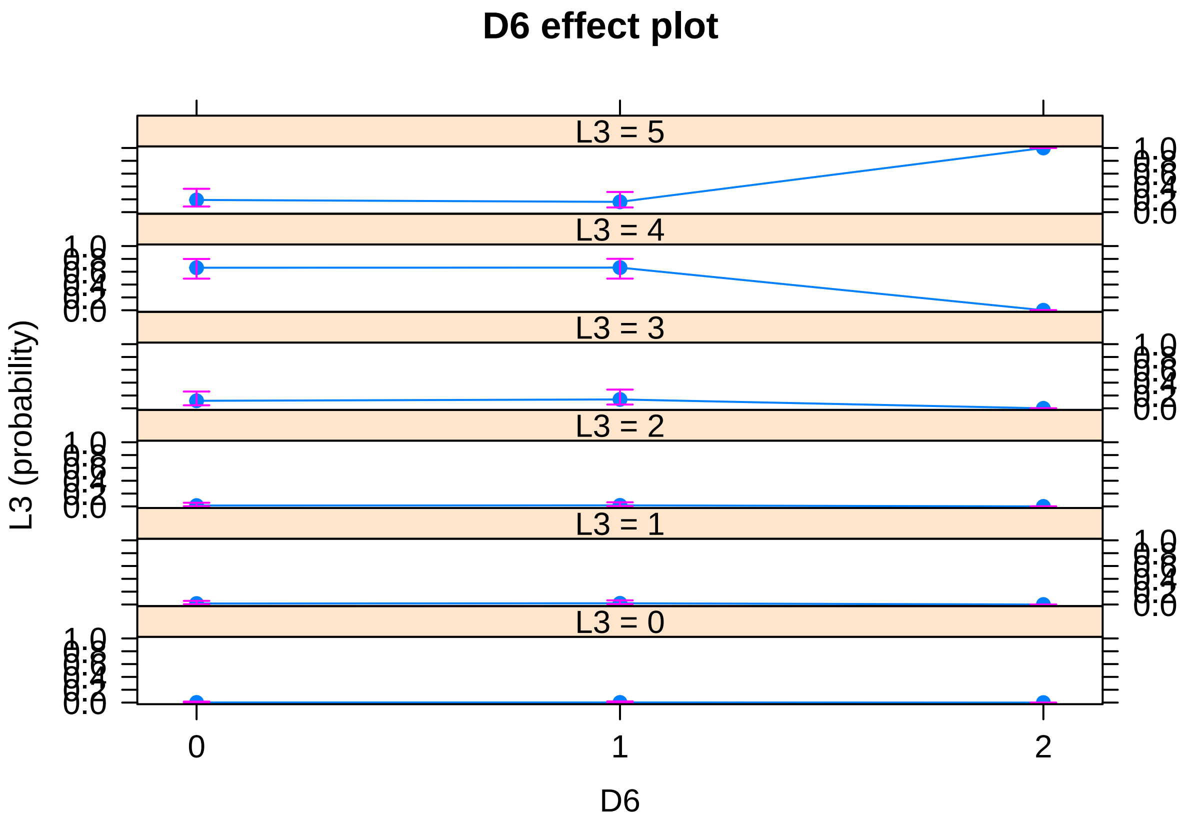
D1 effect plot
L3 = 5
L3 = 4
L3 = 3
L3 = 2
L3 = 1
L3 = 0
L3 (probability)
0.0
0.2
0.4
0.6
0.8
0
1
2
D1

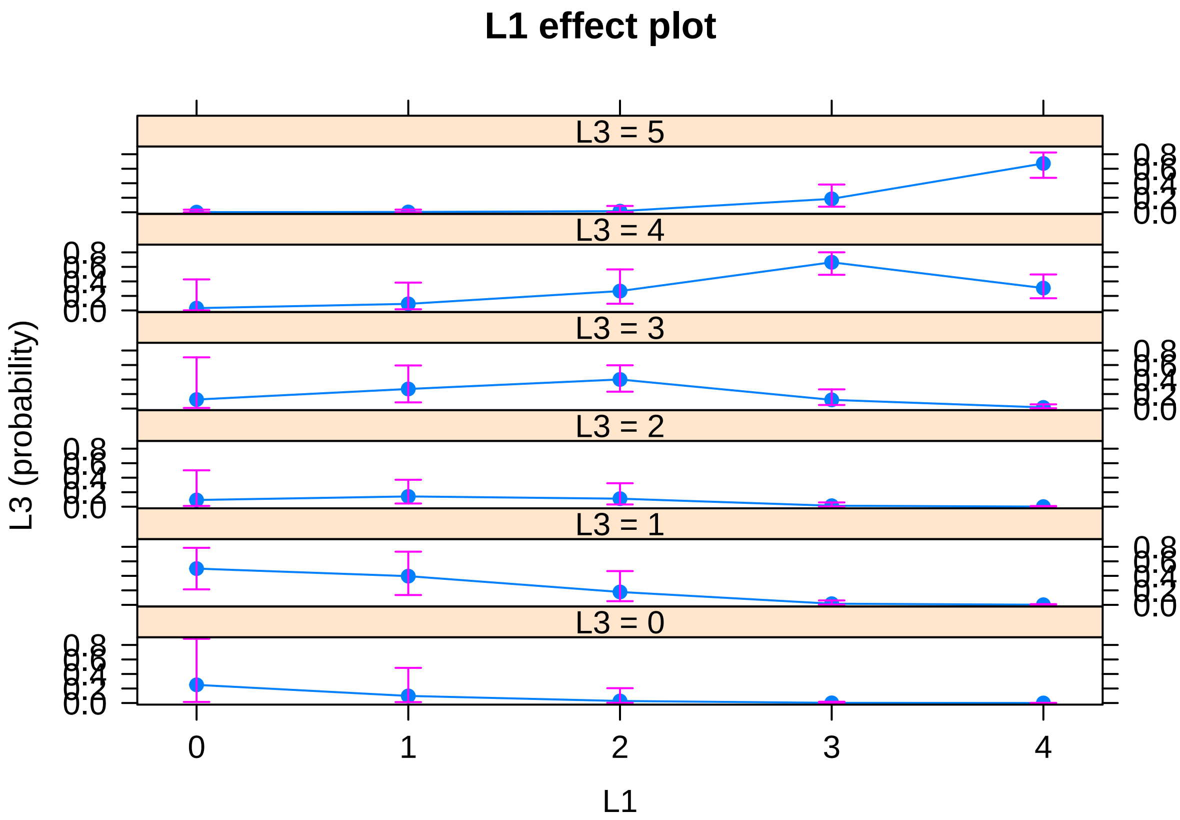


D4 effect plot
L3 = 5
L3 = 4
L3 = 3
L3 = 2
L3 = 1
L3 = 0
0.0
0.2
0.4
0.6
0.8
L3 (probability)
1
2
3
4
5
D4

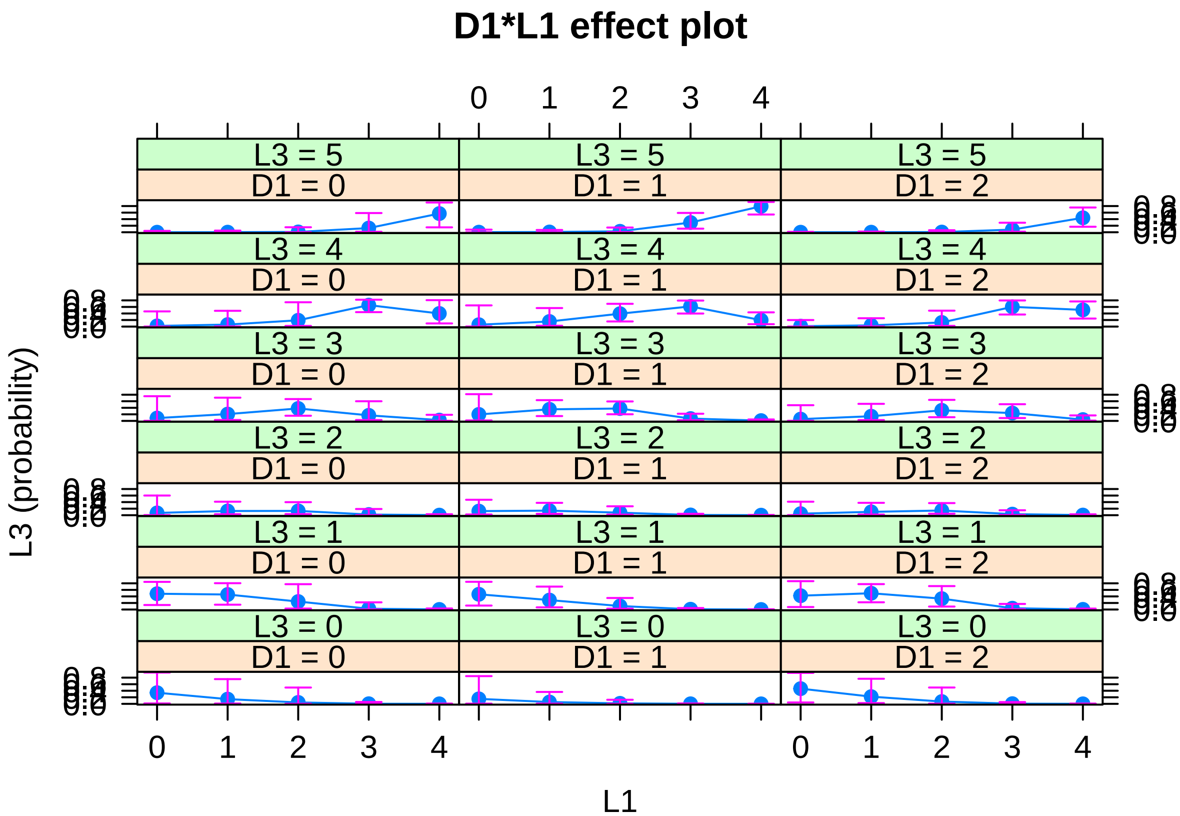
D5 effect plot




D6 effect plot
L3 = 5
L3 = 4
L3 = 3
L3 = 2
L3 = 1
L3 = 0
L3 (probability)
1.0
0.0
0
1
2
D6




```r
# Interaction effects
plot(effects::Effect(focal.predictors = c("D1", "L1"), mod = olr1))
```



Regression Analysis - Effect of Video

Linear Regression

To assess the affect of the video on respondent's view on including climate change curricula, we run a multivariate linear regression where the independent variable is the change (Δ) between there response before and after watching the accompanying video. We must treat this as continuous because $\Delta_i = L3_i - L2_i$. Thus, our regression will be of the form

$$\Delta_i = D1_i + D4_i + D5_i + D6_i + L1_i + \epsilon_i$$

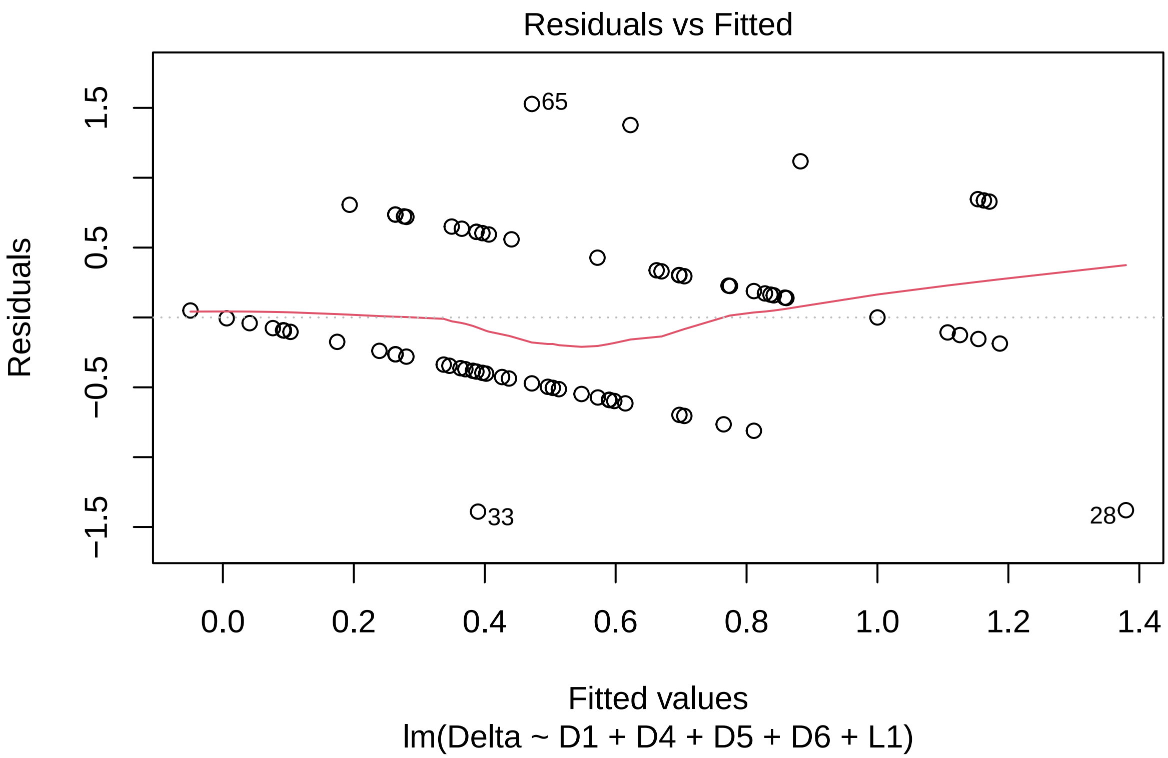
where DX indicates categorical demographic variables and LX indicates ordinal Likert scale variables as previously.

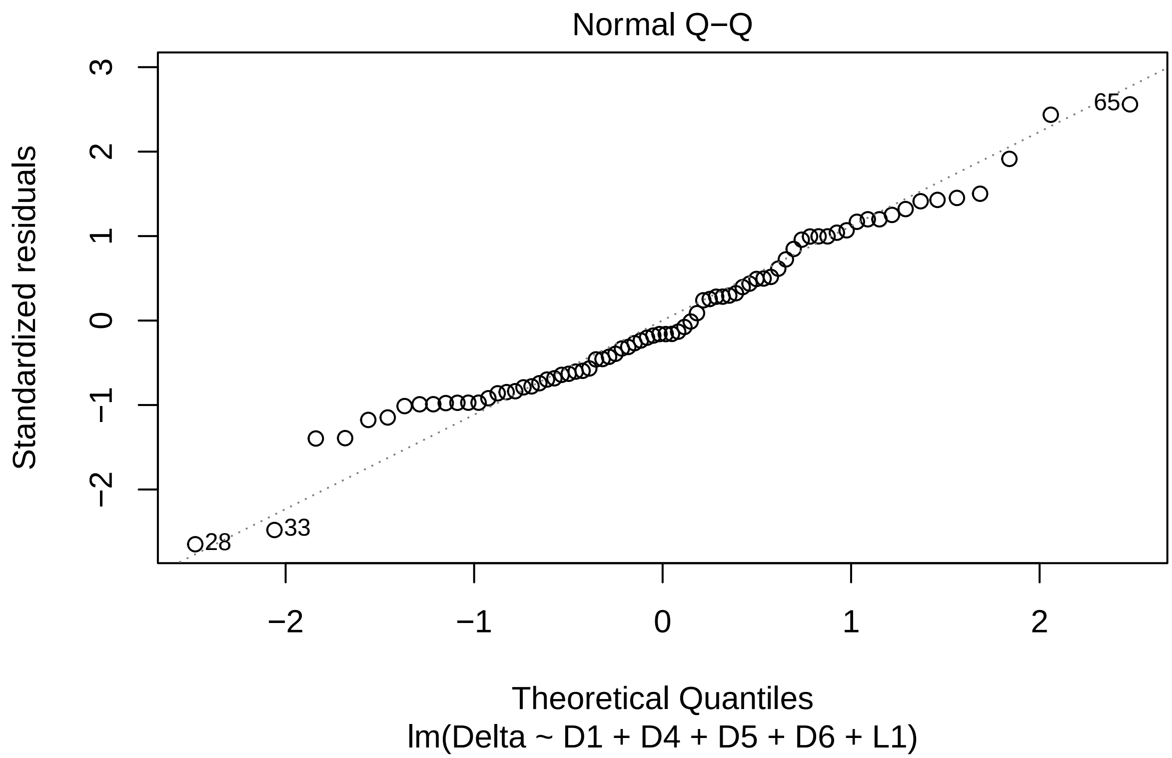
```
# Model 1 - we treat independent variable as continuous
mlr.delta <- lm(Delta ~ D1 + D4 + D5 + D6 + L1, data = reg.data)
mlr.delta.table <- broom::tidy(mlr.delta)
summary(mlr.delta)
```

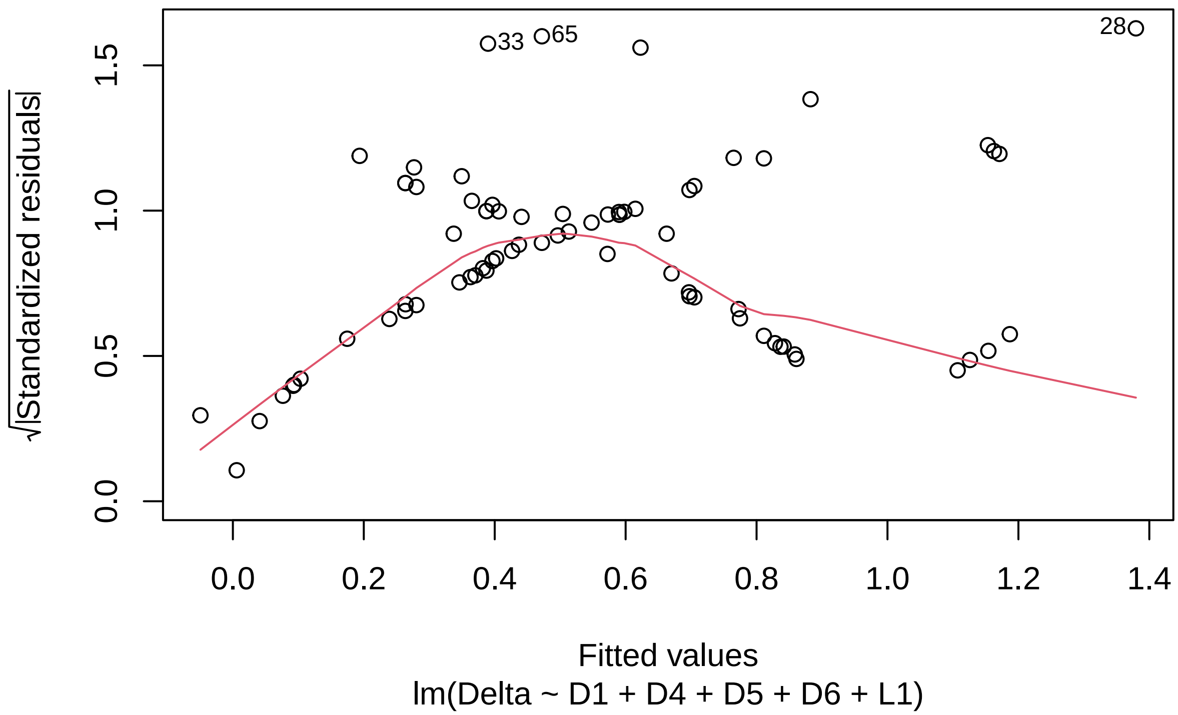
```
##
## Call:
## lm(formula = Delta ~ D1 + D4 + D5 + D6 + L1, data = reg.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -1.38977 -0.40218 -0.09326  0.33736  1.52797
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.57361    0.43316   1.324   0.1904
## D11         -0.40200    0.38460  -1.045   0.3001
## D12          0.06309    0.33811   0.187   0.8526
## D42          0.12593    0.23255   0.541   0.5902
## D43          0.33448    0.20363   1.643   0.1057
## D44          0.57802    0.32889   1.757   0.0839 .
## D45          0.33373    0.25173   1.326   0.1900
## D51          0.10697    0.20579   0.520   0.6051
## D52         -0.64219    0.40170  -1.599   0.1151
## D53          0.11543    0.39043   0.296   0.7685
## D54         -0.77812    0.42603  -1.826   0.0728 .
## D61         -0.02435    0.16137  -0.151   0.8806
## D62          0.71963    0.68140   1.056   0.2952
## L1.L         0.38531    0.33843   1.139   0.2594
## L1.Q        -0.32384    0.32398  -1.000   0.3215
## L1.C         0.05713    0.25249   0.226   0.8218
## L1^4         0.16816    0.22115   0.760   0.4500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6492 on 60 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.2358, Adjusted R-squared:  0.03198
## F-statistic: 1.157 on 16 and 60 DF,  p-value: 0.3279
```

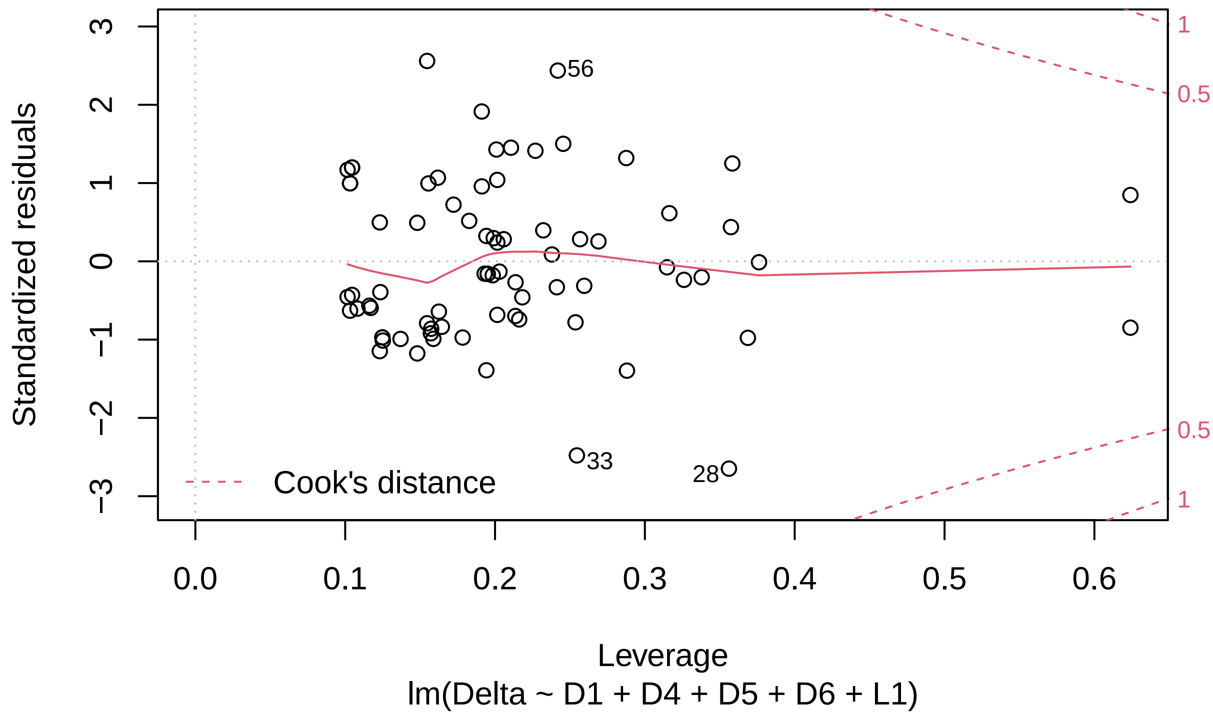
```
plot(mlr.delta, caption = list("Residuals vs Fitted", "Normal Q-Q"))
```


Residuals vs Fitted
65
33
28
1.5
0.5
−0.5
−1.5
Residuals
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Fitted values
lm(Delta ~ D1 + D4 + D5 + D6 + L1)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(Delta ~ D1 + D4 + D5 + D6 + L1)
28
33
65
−2
−1
0
1
2
3



√|Standardized residuals|
0.0
0.5
1.0
1.5
33
65
28
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Fitted values
lm(Delta ~ D1 + D4 + D5 + D6 + L1)




```r
# Run ANOVA
mlr.delta.anova <- anova(mlr.delta)
mlr.delta.anova
```

```
## Analysis of Variance Table
##
## Response: Delta
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## D1         2  0.1279 0.06397  0.1518 0.85951
## D4         4  2.4579 0.61448  1.4579 0.22630
## D5         4  3.6962 0.92405  2.1924 0.08058 .
## D6         2  0.5473 0.27363  0.6492 0.52609
## L1         4  0.9725 0.24313  0.5769 0.68050
## Residuals 60 25.2890 0.42148
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
dev.off()
```

```
## null device
##           1
```