02 - Regression Analysis

Alec Stashevsky

11/27/2021

Build Regression Data

```
# 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

Linear Regression

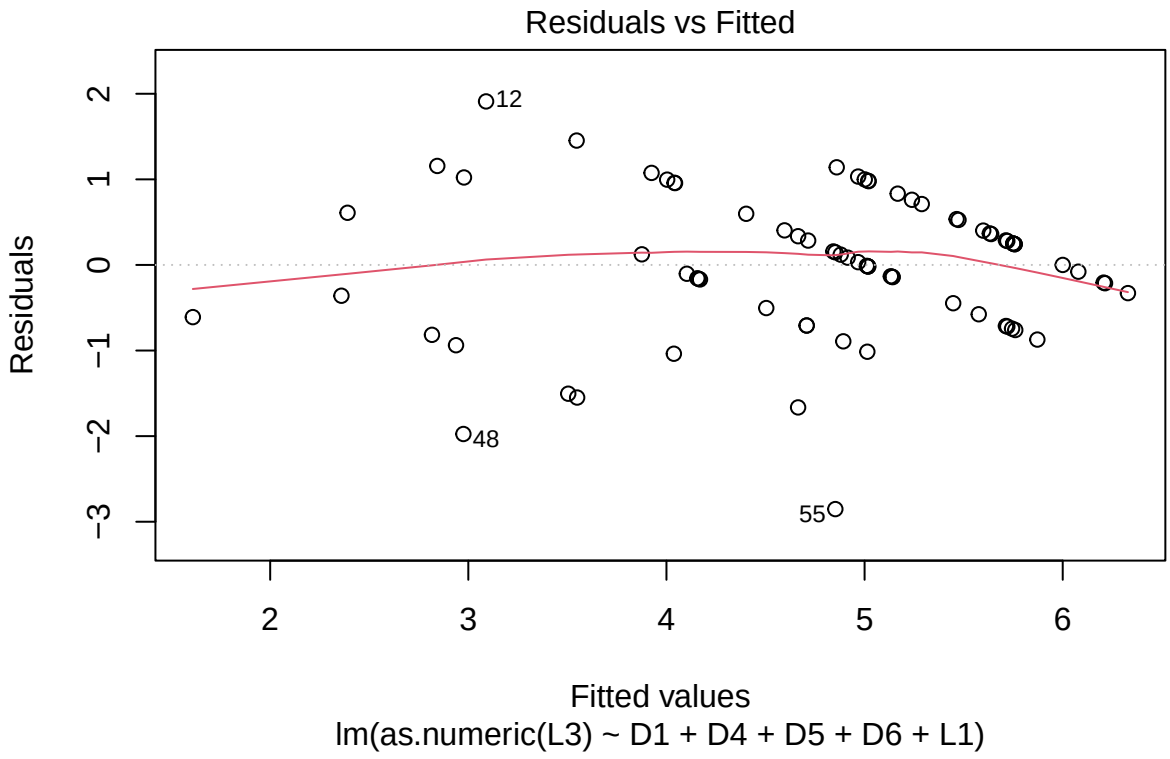
We will first test run a multivariate linear regression where we treat the independent variable as continuous. Our first regression will be of the form

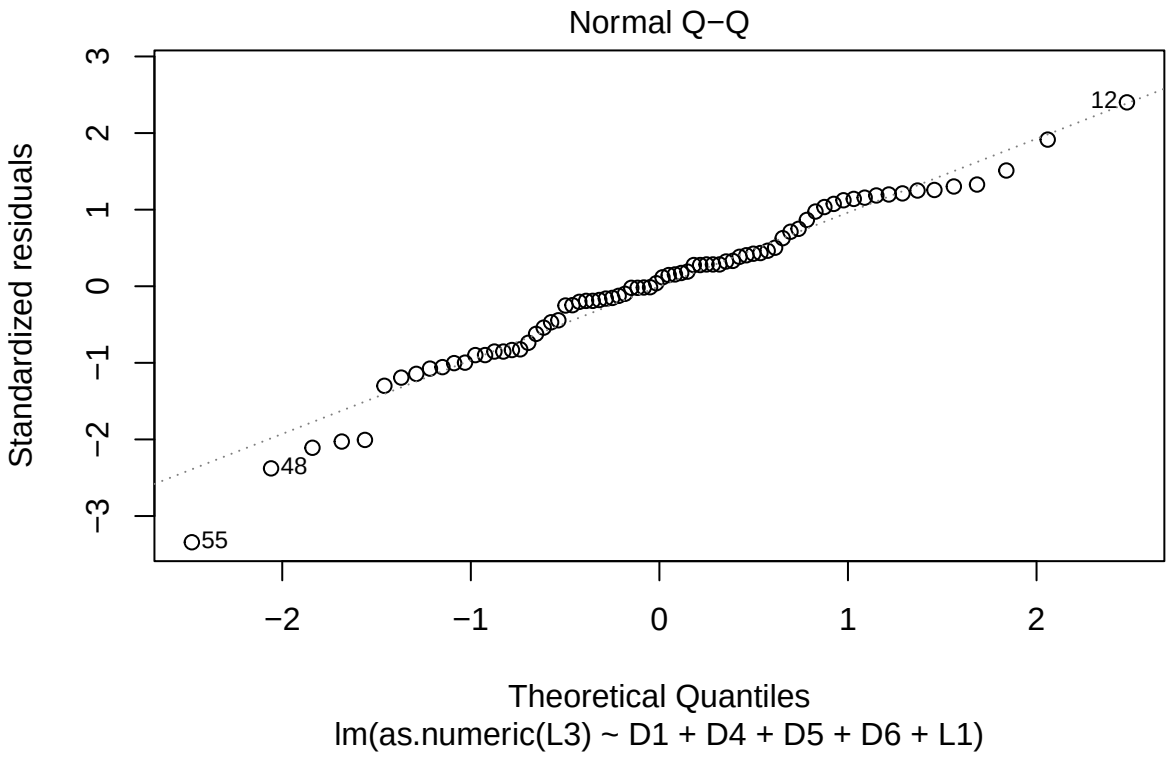
$$L3 = D1 + D4 + D5 + D6 + L1 + \epsilon$$

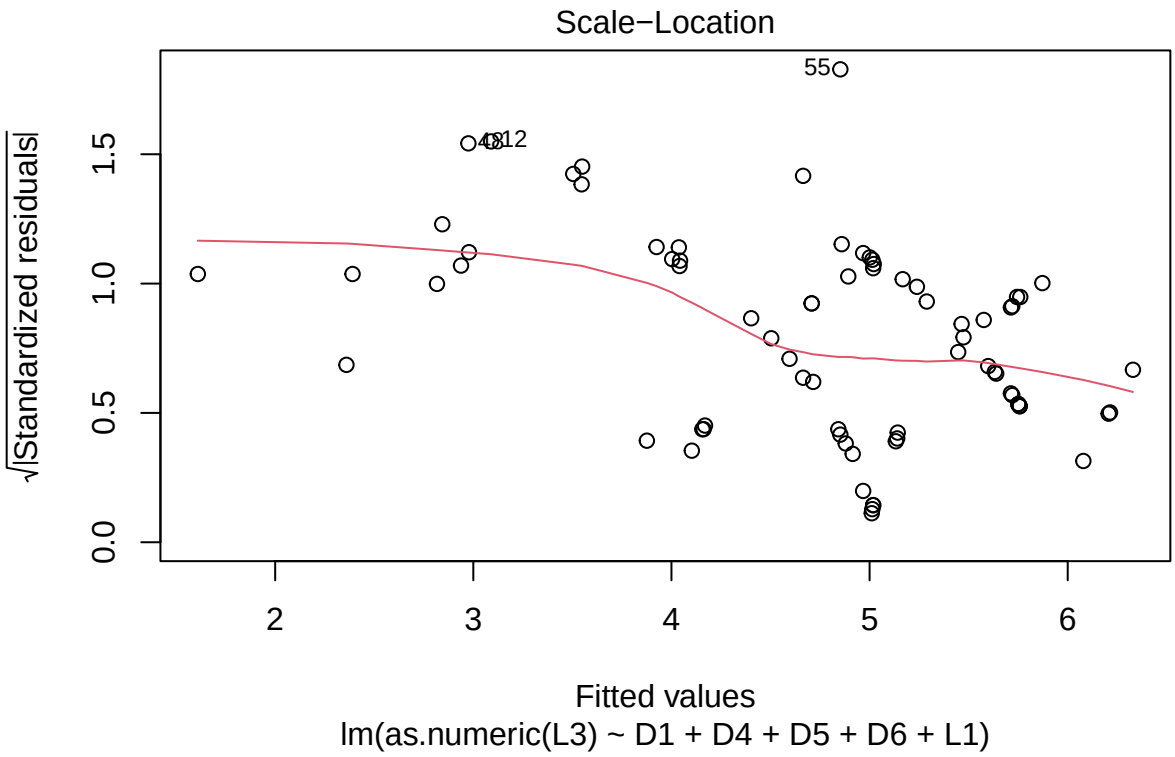
where Dx indicates categorical demographic variables and Lx indicates ordinal Likert scale variables.

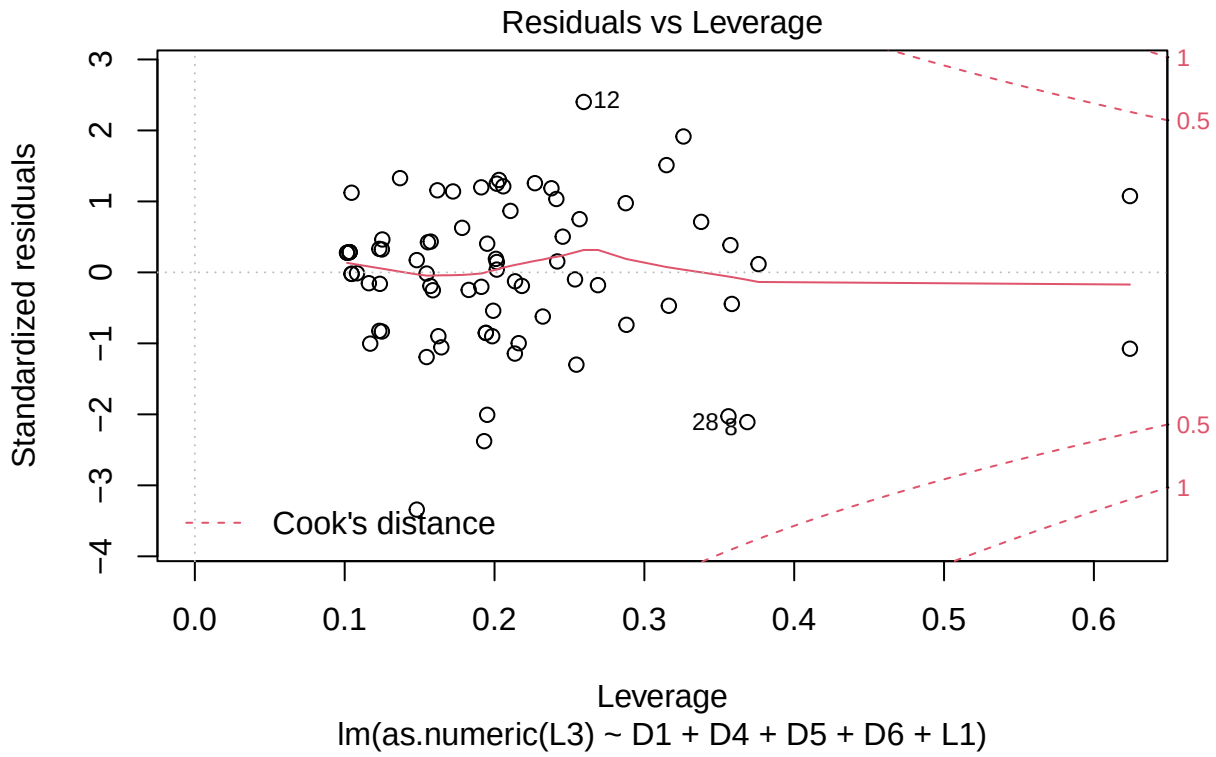
```r
# Model 1 - we treat independent variable as continuous
mlr1 <- lm(as.numeric(L3) ~ D1 + D4 + D5 + D6 + L1, data = reg.data)
mlr1.table <- broom::tidy(mlr1)
plot(mlr1)
```

```
## Warning: not plotting observations with leverage one:
##   49
```



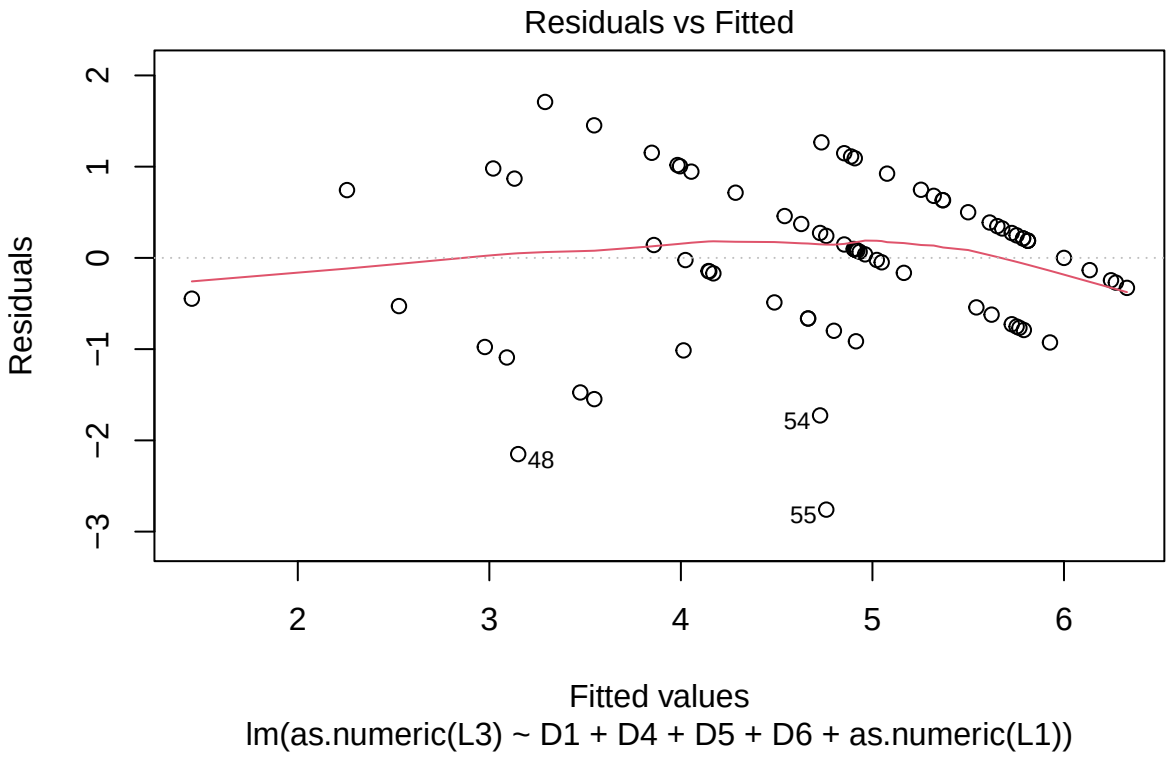

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(as.numeric(L3) ~ D1 + D4 + D5 + D6 + L1)
12
48
55

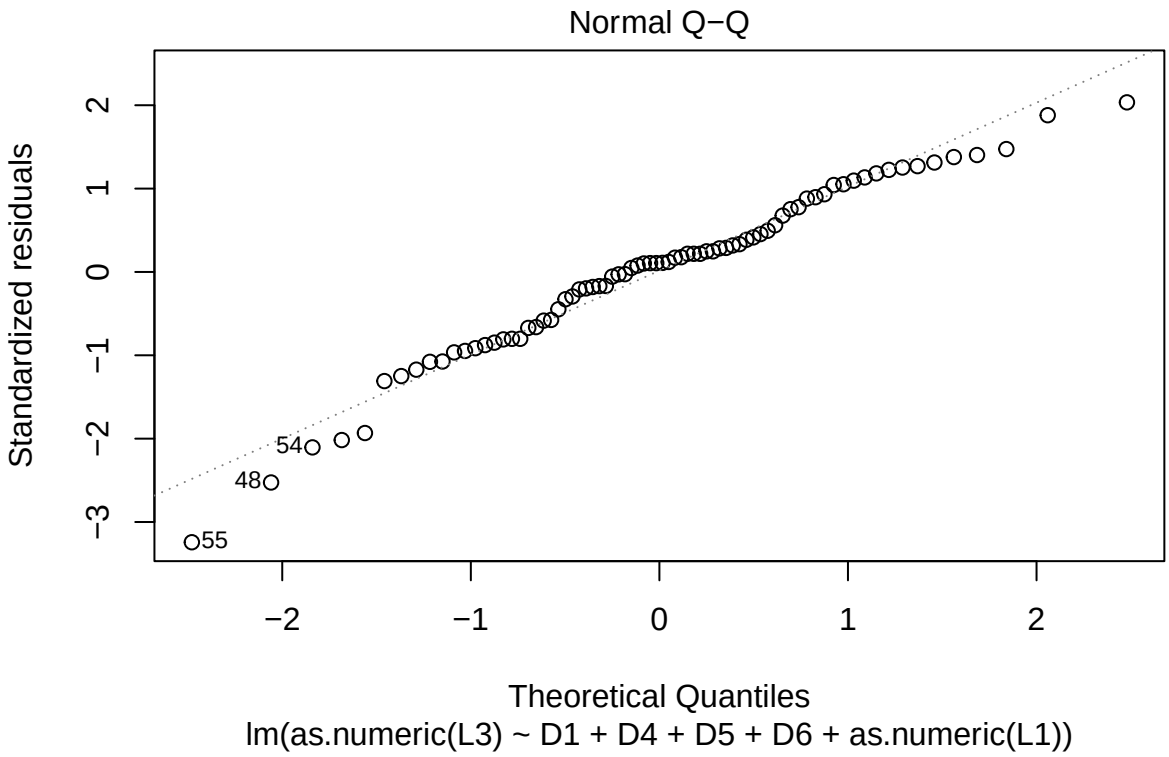

Scale−Location
55
√|Standardized residuals|
0.0
0.5
1.0
1.5
2
3
4
5
6
Fitted values
lm(as.numeric(L3) ~ D1 + D4 + D5 + D6 + L1)

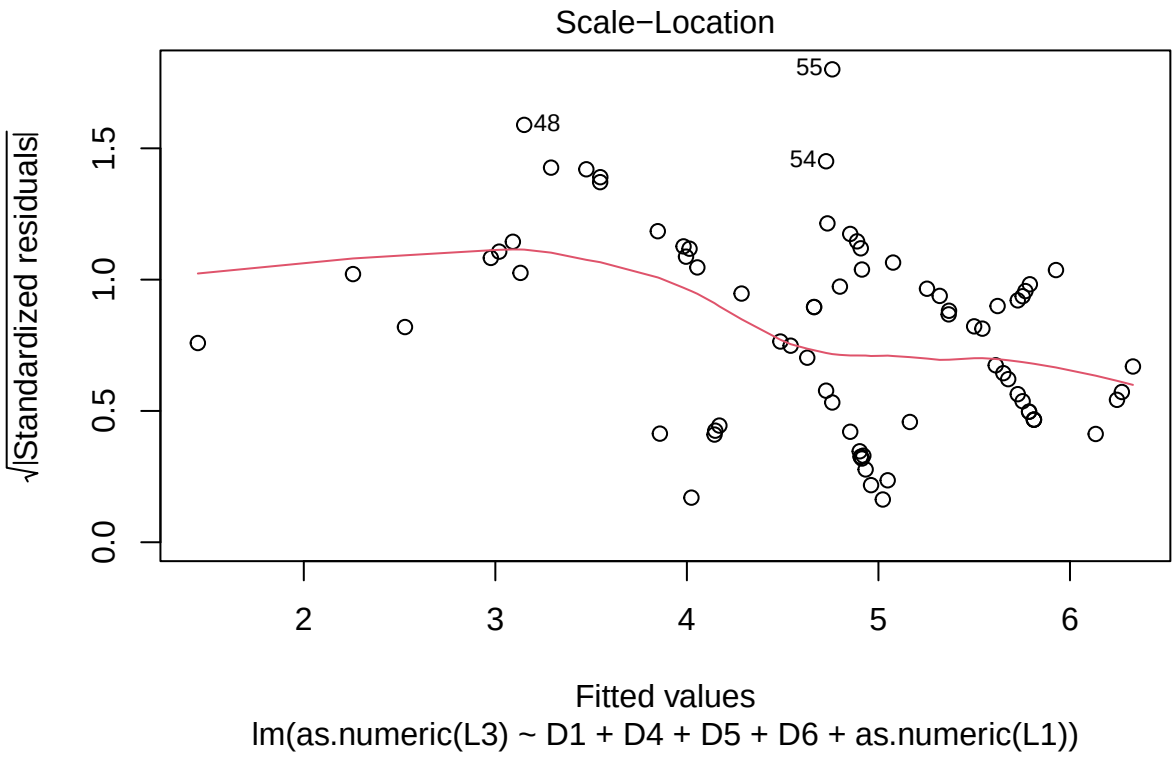


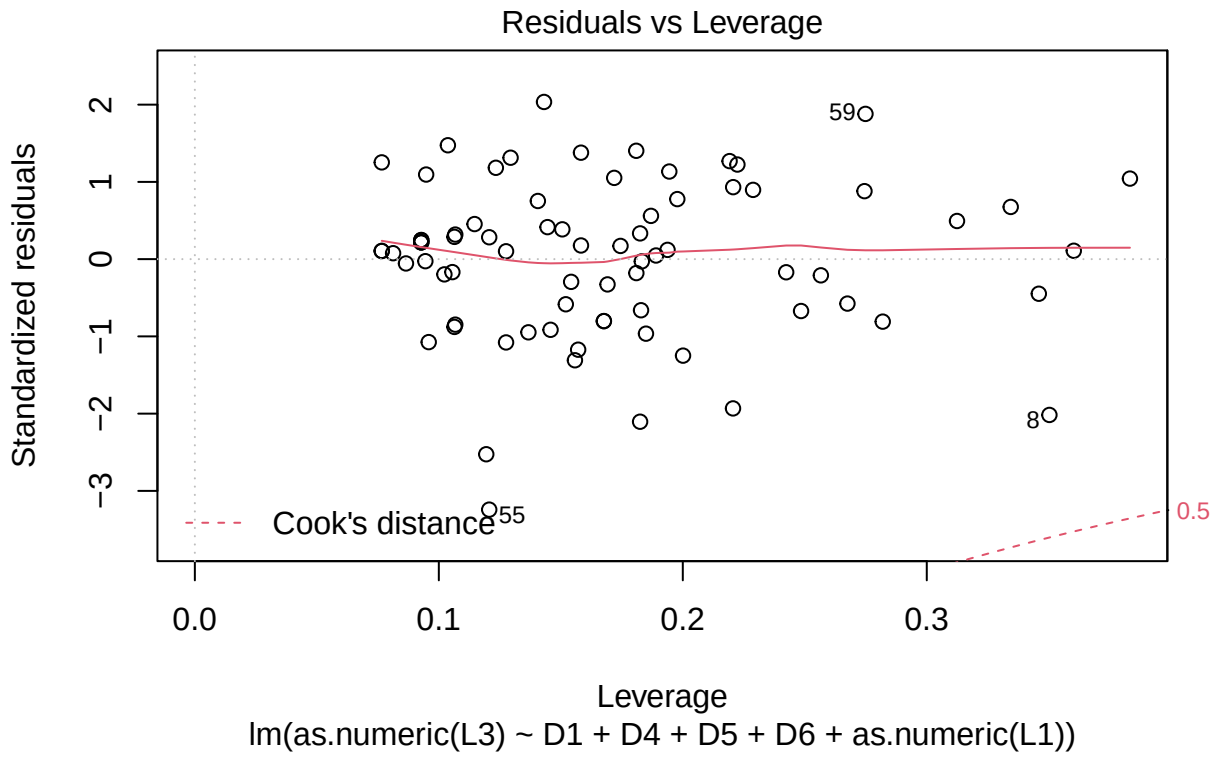

```
# Model 2 - we treat all Likert variables as continuous
mlr2 <- lm(as.numeric(L3) ~ D1 + D4 + D5 + D6 + as.numeric(L1), data = reg.data)
mlr2.table <- broom::tidy(mlr2)
plot(mlr2)
```

```
## Warning: not plotting observations with leverage one:
##   49
```


Residuals vs Fitted
Residuals
2
1
0
−1
−2
−3
2
3
4
5
6
54
48
55
Fitted values
lm(as.numeric(L3) ~ D1 + D4 + D5 + D6 + as.numeric(L1))


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


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



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.

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

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
olr1.table <- broom::tidy(olr1)

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

```
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## --------------------------------------------
## 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

## Warning in brant(olr1): 6678 combinations in table(dv,ivs) do not occur. Because
## of that, the test results might be invalid.
```

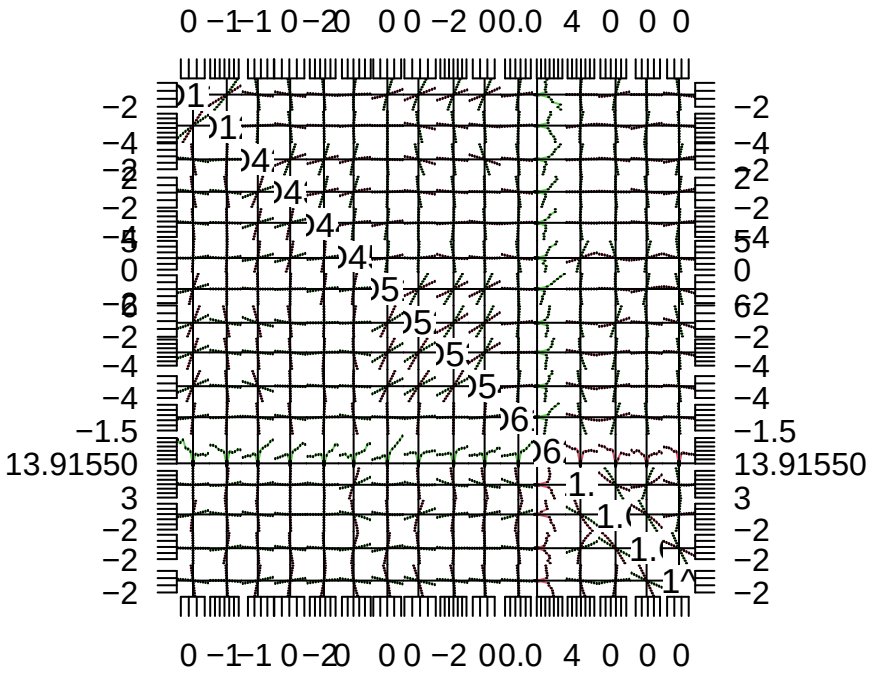
```
# Check goodness of fit
paste("Goodness of fit Chi-sq:", 1-pchisq(deviance(olr1),df.residual(olr1)))
```

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

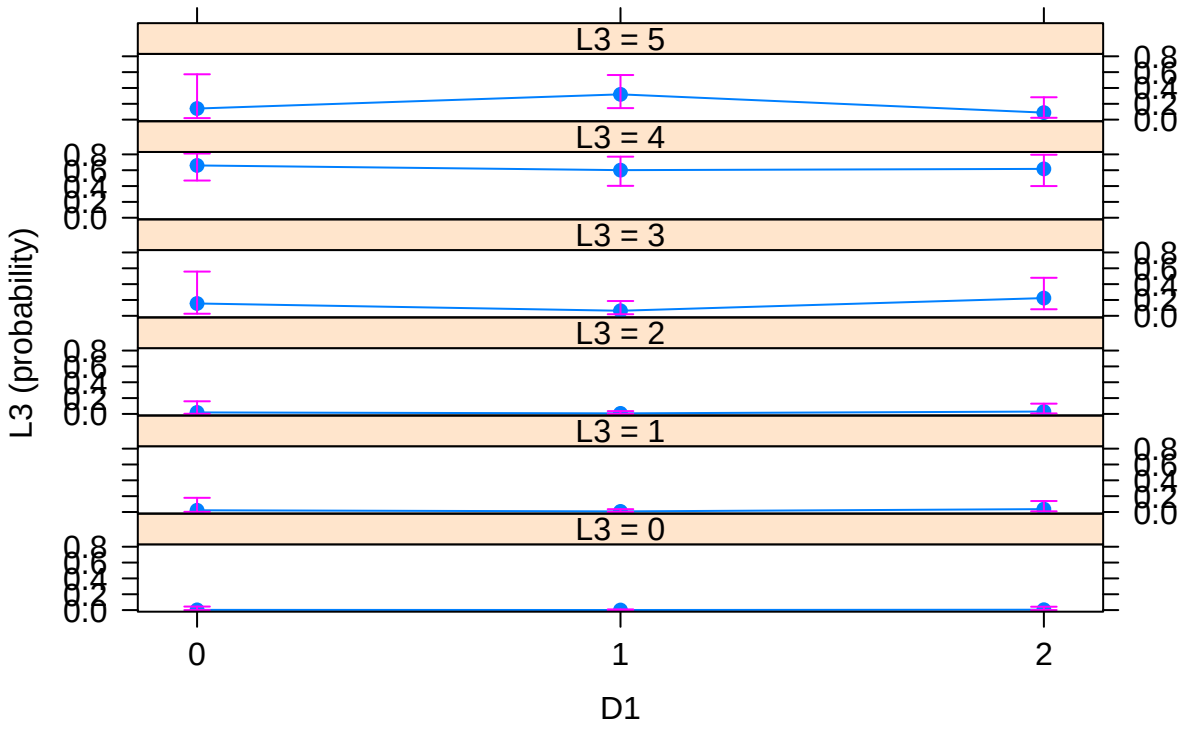
```
# ANOVA
olr1.anova <- broom::tidy(Anova(olr1))
```

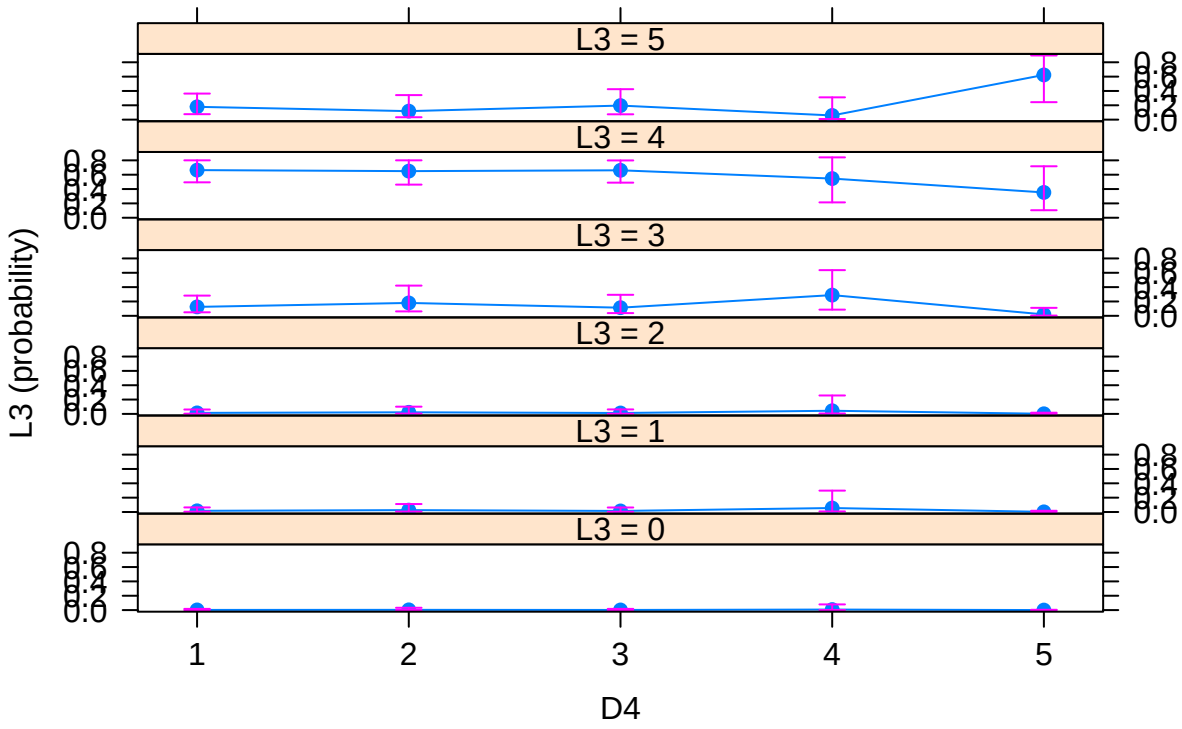
```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

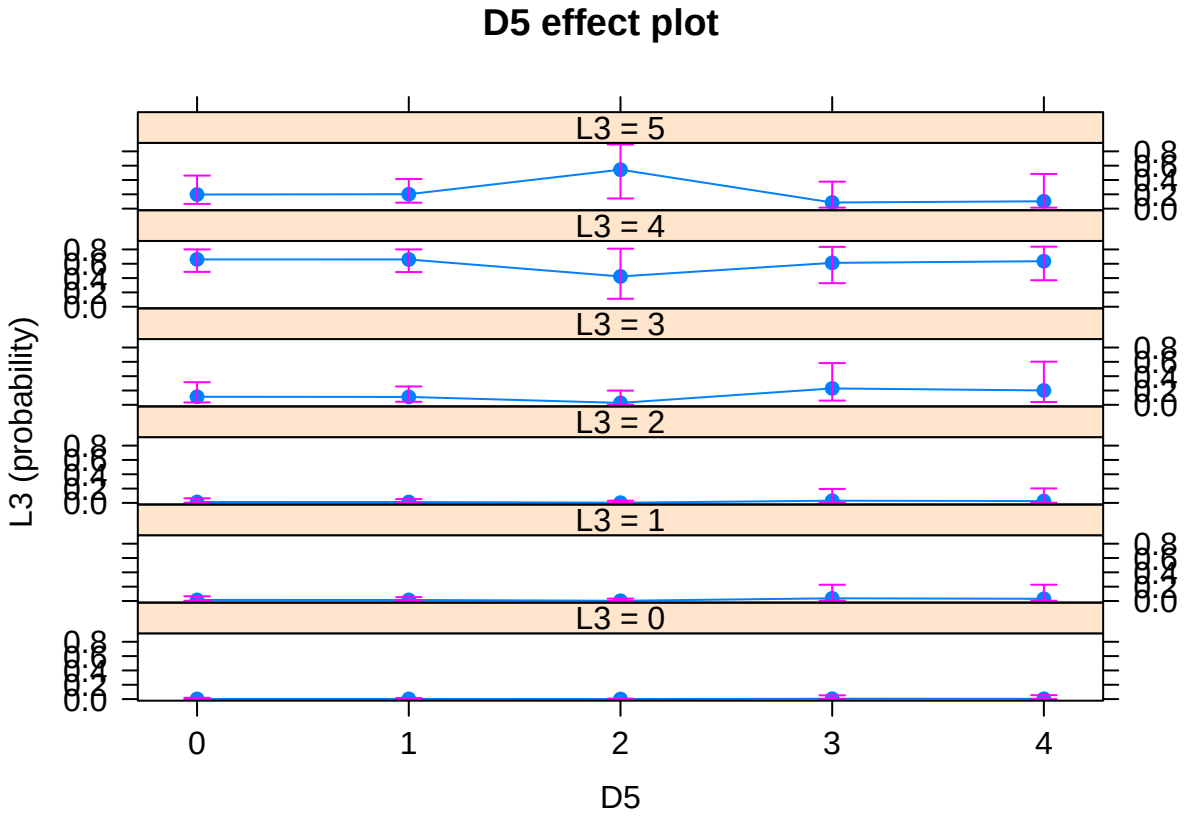
```
# Add p-values to regression table
olr1.table$p.value <- pnorm(abs(olr1.table$statistic), lower.tail = FALSE) * 2

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

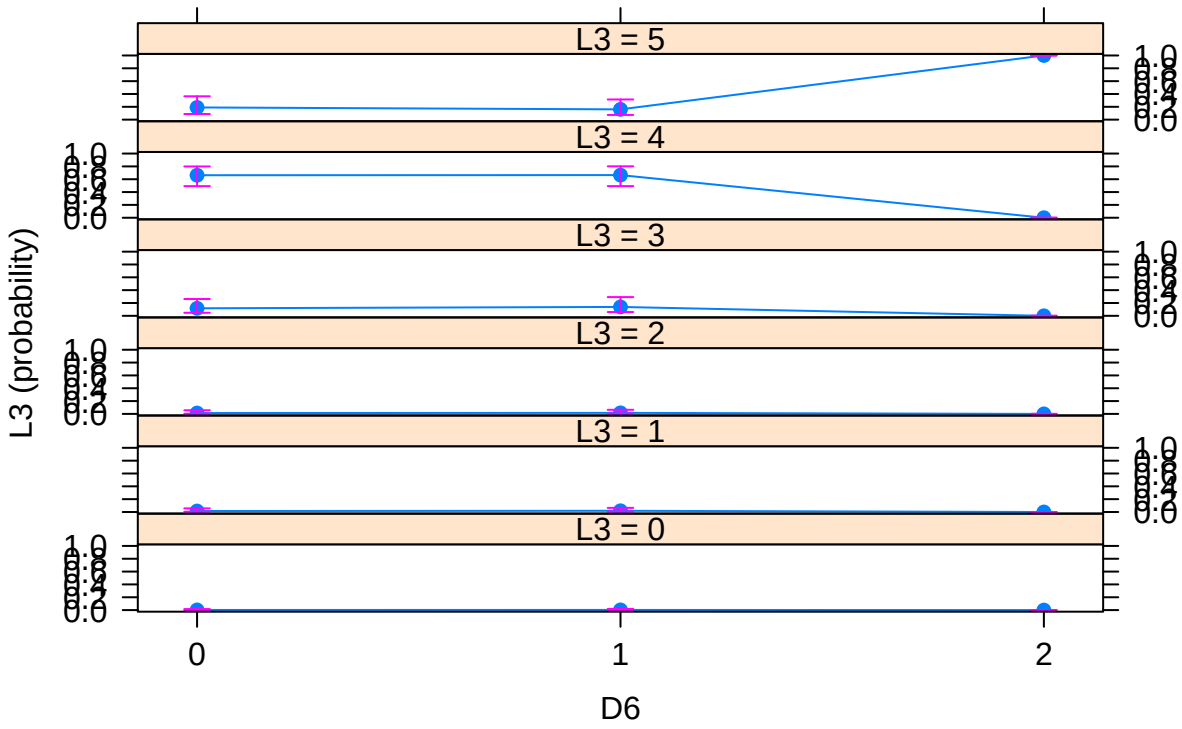


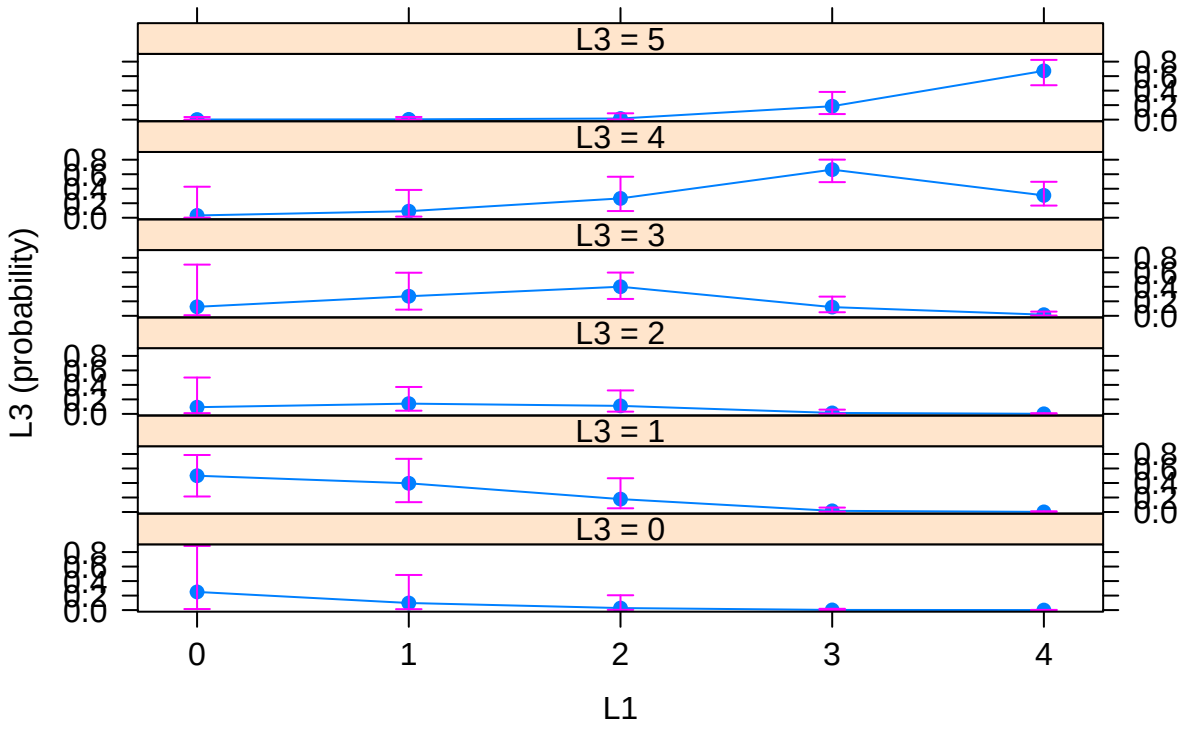

```
predictors <- c("D1", "D4", "D5", "D6", "L1")
for (p in predictors) {
  print(plot(effects::Effect(focal.predictors = c(p), mod = olr1)))
}
```


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

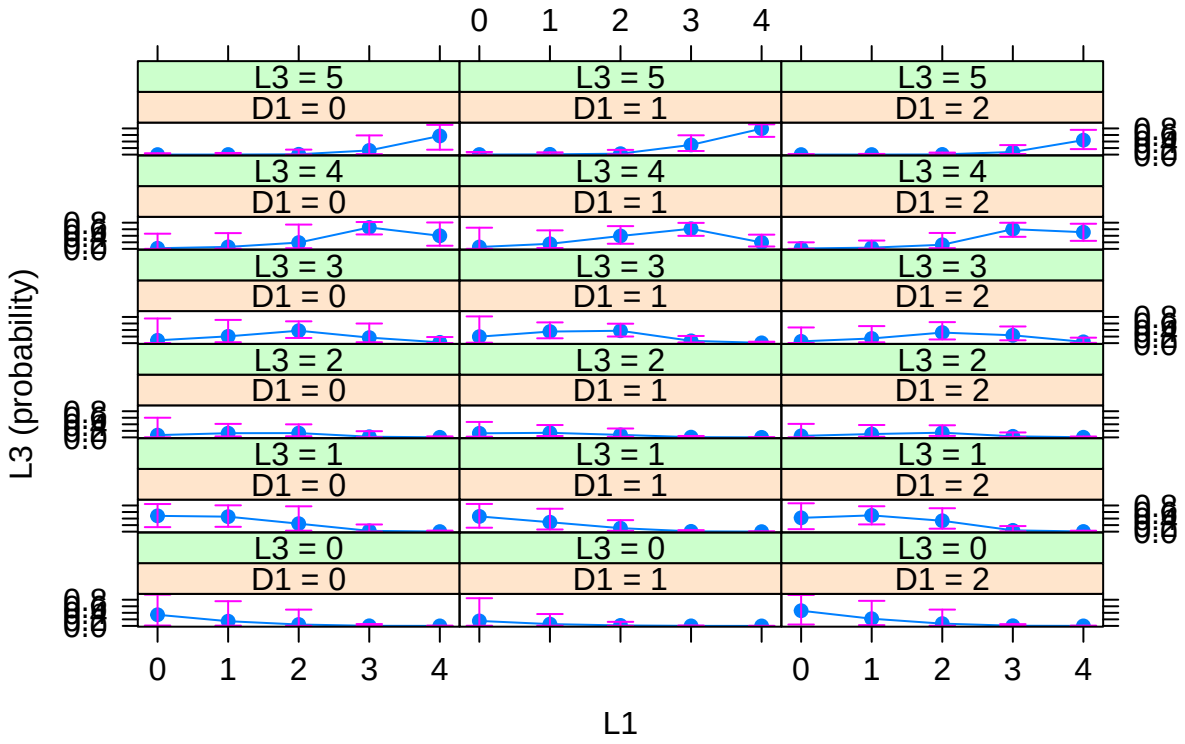

D4 effect plot
L3 = 5
L3 = 4
L3 = 3
L3 = 2
L3 = 1
L3 = 0
L3 (probability)
D4
1
2
3
4
5


D5 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
3
4
D5


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




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




```
dev.off()
```

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