

# Multiple and Logistic Regression in R\_Parallel Slopes

dizhen

4/29/2020

## What if you have two groups?

one continuous variable and one categorical variable: parallel slopes model

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
load("data/mario_kart.RData")
```

```
mario_kart <- mario_kart %>% filter(totalPr<=100)
```

```
# Explore the data  
glimpse(mario_kart)
```

```
## Rows: 141  
## Columns: 12  
## $ ID      <dbl> 150377422259, 260483376854, 320432342985, 280405224677, ...  
## $ duration <int> 3, 7, 3, 3, 1, 3, 1, 1, 3, 7, 1, 1, 1, 1, 7, 7, 3, 3, 1,...  
## $ nBids    <int> 20, 13, 16, 18, 20, 19, 13, 15, 29, 8, 15, 15, 13, 16, 6...  
## $ cond     <chr> "new", "used", "new", "new", "new", "new", "used", "new"...  
## $ startPr  <dbl> 0.99, 0.99, 0.99, 0.99, 0.01, 0.99, 0.01, 1.00, 0.99, 19...  
## $ shipPr   <dbl> 4.00, 3.99, 3.50, 0.00, 0.00, 4.00, 0.00, 2.99, 4.00, 4....  
## $ totalPr  <dbl> 51.55, 37.04, 45.50, 44.00, 71.00, 45.00, 37.02, 53.99, ...  
## $ shipSp   <chr> "standard", "firstClass", "firstClass", "standard", "med...  
## $ sellerRate <int> 1580, 365, 998, 7, 820, 270144, 7284, 4858, 27, 201, 485...  
## $ stockPhoto <chr> "yes", "yes", "no", "yes", "yes", "yes", "yes", "yes", "...  
## $ wheels   <int> 1, 1, 1, 1, 2, 0, 0, 2, 1, 1, 2, 2, 2, 2, 1, 0, 1, 1, 2,...  
## $ title    <chr> "~~ Wii MARIO KART & WHEEL ~ NINTENDO Wii ~ BRAND NE...
```

```
# fit parallel slopes
mod <- lm(totalPr ~ wheels + cond, data = mario_kart)
summary(mod)

##
## Call:
## lm(formula = totalPr ~ wheels + cond, data = mario_kart)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.0078  -3.0754  -0.8254   2.9822  14.1646
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  42.3698     1.0651   39.780 < 2e-16 ***
## wheels        7.2328     0.5419   13.347 < 2e-16 ***
## condused     -5.5848     0.9245   -6.041 1.35e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.887 on 138 degrees of freedom
## Multiple R-squared:  0.7165, Adjusted R-squared:  0.7124
## F-statistic: 174.4 on 2 and 138 DF,  p-value: < 2.2e-16
```

## Visualizing parallel slopes models

Retrieving the coefficients

```
library(broom)
augment(mod)

## # A tibble: 141 x 10
##   totalPr wheels cond .fitted .se.fit .resid .hat .sigma .cooksd .std.resid
##   <dbl>   <int> <chr>   <dbl>   <dbl> <dbl> <dbl> <dbl>   <dbl>   <dbl>
## 1    51.6     1 new     49.6    0.709  1.95  0.0210  4.90  1.16e-3    0.403
## 2    37.0     1 used    44.0    0.547 -6.98  0.0125  4.87  8.71e-3   -1.44
## 3    45.5     1 new     49.6    0.709 -4.10  0.0210  4.89  5.15e-3   -0.848
## 4    44      1 new     49.6    0.709 -5.60  0.0210  4.88  9.61e-3   -1.16
## 5    71      2 new     56.8    0.676 14.2   0.0192  4.75  5.57e-2    2.93
## 6    45      0 new     42.4    1.07   2.63  0.0475  4.90  5.05e-3    0.551
## 7    37.0     0 used    36.8    0.707  0.235  0.0209  4.91  1.68e-5    0.0486
## 8    54.0     2 new     56.8    0.676 -2.85  0.0192  4.90  2.25e-3   -0.588
## 9    47      1 used    44.0    0.547  2.98  0.0125  4.90  1.59e-3    0.614
## 10   50      1 used    44.0    0.547  5.98  0.0125  4.88  6.40e-3    1.23
## # ... with 131 more rows
```

Parallel lines on the scatterplot

```
data_space +
  geom_line(data = augment(mod), aes(y = .fitted, color = factor.year))
```

```

library(ggplot2)

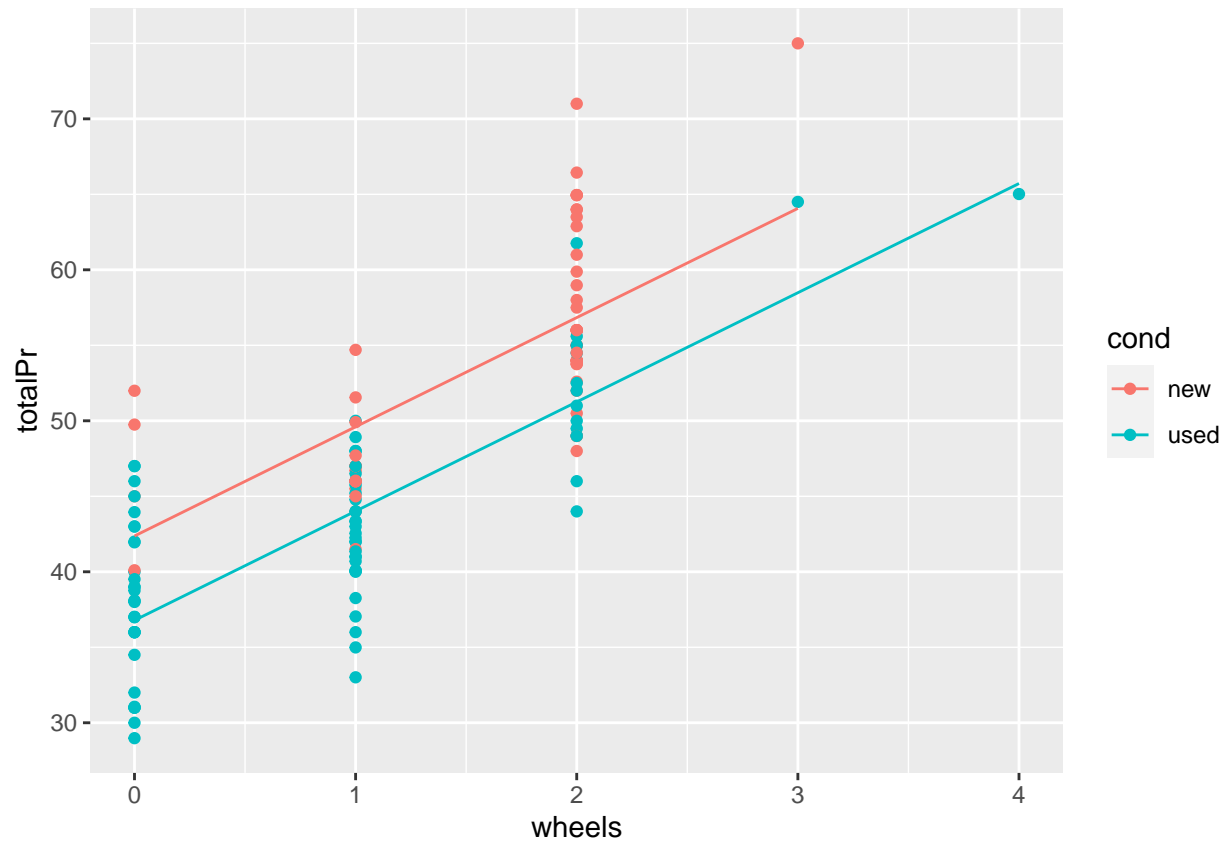
# Augment the model
augmented_mod <- augment(mod)
glimpse(augmented_mod)

## Rows: 141
## Columns: 10
## $ totalPr    <dbl> 51.55, 37.04, 45.50, 44.00, 71.00, 45.00, 37.02, 53.99, ...
## $ wheels     <int> 1, 1, 1, 1, 2, 0, 0, 2, 1, 1, 2, 2, 2, 2, 1, 0, 1, 1, 2,...
## $ cond       <chr> "new", "used", "new", "new", "new", "new", "used", "new"...
## $ .fitted     <dbl> 49.60260, 44.01777, 49.60260, 49.60260, 56.83544, 42.369...
## $ .se.fit     <dbl> 0.7087865, 0.5465195, 0.7087865, 0.7087865, 0.6764502, 1...
## $ .resid      <dbl> 1.9473995, -6.9777674, -4.1026005, -5.6026005, 14.164559...
## $ .hat        <dbl> 0.02103158, 0.01250410, 0.02103158, 0.02103158, 0.019156...
## $ .sigma      <dbl> 4.902339, 4.868399, 4.892414, 4.881308, 4.750591, 4.8998...
## $ .cooksd     <dbl> 1.161354e-03, 8.712334e-03, 5.154337e-03, 9.612441e-03, ...
## $ .std.resid  <dbl> 0.40270893, -1.43671086, -0.84838977, -1.15857953, 2.926...

# scatterplot, with color
data_space <- ggplot(augmented_mod, aes(x = wheels, y = totalPr, color = cond)) +
  geom_point()

# single call to geom_line()
data_space +
  geom_line(aes(y = .fitted))

```



## Interpreting parallel slopes coefficients

### Three ways to describe a model

```
library(UsingR)
```

```
## Loading required package: MASS
```

```
##
```

```
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
```

```
## select
```

```
## Loading required package: HistData
```

```
## Loading required package: Hmisc
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':
##
##      src, summarize

## The following objects are masked from 'package:base':
##
##      format.pval, units

##
## Attaching package: 'UsingR'

## The following object is masked from 'package:survival':
##
##      cancer
```

```
data(babies)
# build model
lm(wt1 ~ age + parity, data = babies)
```

```
##
## Call:
## lm(formula = wt1 ~ age + parity, data = babies)
##
## Coefficients:
## (Intercept)      age      parity
##    120.227      1.022      2.985
```

```
# build model
lm(wt1 ~ gestation + smoke, data = babies)
```

```
##
## Call:
## lm(formula = wt1 ~ gestation + smoke, data = babies)
##
## Coefficients:
## (Intercept)  gestation      smoke
##    127.14132    0.09772   -1.38337
```