Dummie Variables

Dummy variables are included in models to include explanatory variables that are not of metric scale (but ordinal or nominal), e.g. whether a certain event has taken place or not. Dummy variables have values of zero or one, i.e. there are two possible outcomes.

Intercept dummies

Suppose you want to know what factors determine the GPA of highschool students across the United States. You might want to know the effect of having a photographic memory on the GPA. Hence, you create an intercept dummy that is zero for people without a photographic memory and one for people who have a photographic memory. This is a typical example of an intercept dummy. We assume that a certain variable only causes a shift in the intercept without changing anything else.

Slope dummies

You might include the number of hours students spent on studying in your regression modelto explain the GPA.However, people with a photographic memory will have a very different relationship betweenthe number of hours spent on studying and the GPA compared to students without a photographic memory. Therefore, we would include an interaction term between the hours of studying and our dummy variable on top of the hours of studying in our regression model.Thereby, we will generate one coefficient that captures the effect of studying on the GPA for students without a photographic memory, and a second coefficient that captures the additional payoff of one hour studying for students with a photographic memory. This is what we refer to as a slope dummy.

Intercept and slope dummies

For the above example (and in many other applications), it would be wise to include the dummy variables once as a slope dummy and once as an intercept dummy. The students with a photographic memory will then have a different intercept (generally better GPA) and a different relationship of time spent on studying compared to other students.Hence, we have a slope and intercept dummy.

Several outcomes in one variable

We might also have cases where we have a variable with several outcomes. In our example, we would control for the city that the students are living in. We have to create as many dummies as there are cities in our sample minus one (!!!) to avoid perfect multi-collinearity across the dummies and the intercept. Alternatively, we

could also estimate the model with dummies for all cities (NOT minus one) and exclude an intercept. We would obtain city-specific intercepts then.

Eaxample

We first import the data “MarketPower.xlsx”

```
library(readxl)
mapo <- read_excel("D:/data/Empirical Research/5 Dummy
Variables/MarketPower.xlsx")
head(mapo)
```

```
## # A tibble: 6 × 24
##      id  year FixedassetsthEUR StockthEUR TotalassetsthEUR
##   <dbl> <dbl>            <dbl>      <dbl>            <dbl>
## 1     2  2014          307300.     77463.          637867.
## 2     3  2010          172835.     33226.          552348.
## 3     3  2014          207326.     31557.          479163.
## 4     3  2016          195786.     28497.          503652.
## 5     3  2017          187857.     30258.          485474.
## 6     4  2009          222275.     47654.          413396.
## # ℹ 19 more variables: ShareholdersfundsthEUR <dbl>, RevenuethEUR
<dbl>,
## #   SalesthEUR <dbl>, PLbeforetaxthEUR <dbl>, TaxationthEUR <dbl>,
## #   NetincomethEUR <dbl>, CostsofemployeesthEUR <dbl>,
InterestpaidthEUR <dbl>,
## #   DebtorsthEUR <dbl>, Numberofemployees <dbl>, ExportrevenuethEUR
<dbl>,
## #   MaterialcoststhEUR <dbl>, age <dbl>, FC <dbl>, FCR <dbl>, ROA
<dbl>,
## #   eqshare <dbl>, RevGR <dbl>, Markup <dbl>
```

```
summary(mapo)
```

```
##        id             year      FixedassetsthEUR      StockthEUR
##  Min.   :  2.0   Min.   :2009   Min.   :     0.01   Min.   :
0.64
##  1st Qu.: 99.0   1st Qu.:2011   1st Qu.:   369.81   1st Qu.:
159.22
##  Median :211.5   Median :2013   Median :  1094.67   Median :
570.29
##  Mean   :266.9   Mean   :2013   Mean   : 11037.69   Mean   :
4575.00
##  3rd Qu.:360.0   3rd Qu.:2015   3rd Qu.:  4157.77   3rd Qu.:
2445.29
##  Max.   :854.0   Max.   :2017   Max.   :307300.03   Max.
:115594.53
##
##  TotalassetsthEUR   ShareholdersfundsthEUR  RevenuethEUR
##  Min.   :    57.5   Min.   :     0.77      Min.   :     87.8
##  1st Qu.:  1347.7   1st Qu.:   380.46      1st Qu.:   2374.9
```

```
##  Median :  4023.6   Median :  1226.03      Median :   6471.1
##  Mean   : 25972.1   Mean   :  9742.84      Mean   :  60579.3
##  3rd Qu.: 15302.3   3rd Qu.:  5288.93      3rd Qu.:  38280.1
##  Max.   :637866.8   Max.   :221237.52      Max.   :1904158.2
##
##    SalesthEUR        PLbeforetaxthEUR    TaxationthEUR
NetincomethEUR
##  Min.   :     86.3   Min.   :-40711.22   Min.   :-3245.7   Min.   :-
74575.49
##  1st Qu.:   2213.1   1st Qu.:    13.79   1st Qu.:    0.0   1st Qu.:
9.99
##  Median :   6035.1   Median :    94.86   Median :    2.7   Median :
82.45
##  Mean   :  55727.4   Mean   :  1376.51   Mean   :  387.9   Mean   :
733.29
##  3rd Qu.:  35189.9   3rd Qu.:   426.05   3rd Qu.:   92.2   3rd Qu.:
315.87
##  Max.   :1678914.9   Max.   :138773.03   Max.   :36945.3   Max.   :
87741.72
##
##  CostsofemployeesthEUR InterestpaidthEUR   DebtorsthEUR
Numberofemployees
##  Min.   :    12.16     Min.   :  -0.976   Min.   :     0.11   Min.
:   1.0
##  1st Qu.:   280.30     1st Qu.:   4.236   1st Qu.:   310.76   1st
Qu.:   6.0
##  Median :   975.02     Median :  15.752   Median :   835.96   Median
:  21.0
##  Mean   :  5667.68     Mean   : 131.763   Mean   :  5488.00   Mean
: 114.3
##  3rd Qu.:  3407.14     3rd Qu.:  66.089   3rd Qu.:  3656.96   3rd
Qu.:  81.0
##  Max.   :195330.71     Max.   :5601.592   Max.   :186865.22   Max.
:4615.0
##
##  ExportrevenuethEUR MaterialcoststhEUR       age               FC
##  Min.   :    -3.8   Min.   :      2.7   Min.   :  1.00   Min.   :
14.8
##  1st Qu.:     0.0   1st Qu.:   1447.4   1st Qu.: 14.00   1st Qu.:
338.2
##  Median :     0.0   Median :   3797.9   Median : 25.00   Median :
1175.1
##  Mean   : 10783.0   Mean   :  40980.5   Mean   : 33.07   Mean   :
12422.9
##  3rd Qu.:  1246.5   3rd Qu.:  23534.2   3rd Qu.: 50.00   3rd Qu.:
5747.4
##  Max.   :366990.4   Max.   :1513689.7   Max.   :117.00   Max.
:695486.8
##  NA's   :1
##       FCR                ROA                eqshare             RevGR
```

```
##  Min.   : 0.2189   Min.   :-0.446323   Min.   : 0.01582   Min.   :-
77.983
##  1st Qu.:12.6032   1st Qu.: 0.005531   1st Qu.:27.58443   1st Qu.: -
3.026
##  Median :18.1771   Median : 0.038869   Median :40.34883   Median :
4.881
##  Mean   :19.9883   Mean   : 0.048962   Mean   :40.14749   Mean   :
6.194
##  3rd Qu.:25.4689   3rd Qu.: 0.080204   3rd Qu.:54.47316   3rd Qu.:
12.217
##  Max.   :82.0954   Max.   : 0.522538   Max.   :91.85318   Max.
:409.126
##
##      Markup
##  Min.   :-11.496
##  1st Qu.:  2.499
##  Median :  3.000
##  Mean   :  2.607
##  3rd Qu.:  3.476
##  Max.   : 27.018
##
```

We now want to know whether we do observe year-specific effects in markup that deviate from a linear trend. To that direction let us first create the year dummies:

```
library(fastDummies)
library(recipes)
mapo <- dummy_cols(mapo, select_columns = 'year')
```

Now, we can run our linear model including a time trend

```
OLSbase = lm(Markup~RevGR+eqshare+FCR+age+TotalassetsthEUR+year,
dat=mapo)
summary(OLSbase)
```

```
##
## Call:
## lm(formula = Markup ~ RevGR + eqshare + FCR + age + TotalassetsthEUR
+
##     year, data = mapo)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.2355  -0.2517   0.3769   1.0268  24.0547
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.083e+02  5.622e+01  -3.706 0.000219 ***
## RevGR             1.791e-03  3.028e-03   0.592 0.554154
## eqshare          -8.032e-04  3.648e-03  -0.220 0.825784
## FCR               1.798e-02  6.778e-03   2.653 0.008061 **
```

```
## age                7.178e-03  2.946e-03   2.437 0.014943 *
## TotalassetsthEUR   5.104e-07  1.065e-06   0.479 0.631731
## year               1.045e-01  2.793e-02   3.742 0.000190 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.632 on 1381 degrees of freedom
## Multiple R-squared:  0.01853,    Adjusted R-squared:  0.01427
## F-statistic: 4.345 on 6 and 1381 DF,  p-value: 0.0002357
```

As a following step we check what we observe when we include our time dummies

```
OLSdum =
lm(Markup~RevGR+eqshare+FCR+age+TotalassetsthEUR+year_2010+year_2011

+year_2012+year_2013+year_2014+year_2015+year_2016+year_2017,dat=mapo)
summary(OLSdum)

##
## Call:
## lm(formula = Markup ~ RevGR + eqshare + FCR + age + TotalassetsthEUR
+
##     year_2010 + year_2011 + year_2012 + year_2013 + year_2014 +
##     year_2015 + year_2016 + year_2017, data = mapo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9608 -0.6057 -0.1581  0.3522 22.4378
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.555e+00  1.696e-01  15.068  < 2e-16 ***
## RevGR             1.159e-03  1.665e-03   0.696    0.487
## eqshare           2.725e-03  1.928e-03   1.414    0.158
## FCR               2.647e-02  3.664e-03   7.224 8.32e-13 ***
## age               1.912e-03  1.556e-03   1.229    0.219
## TotalassetsthEUR -1.205e-07  5.623e-07  -0.214    0.830
## year_2010         1.749e-01  1.646e-01   1.062    0.288
## year_2011        -3.004e-02  1.693e-01  -0.177    0.859
## year_2012        -8.098e+00  1.789e-01 -45.268  < 2e-16 ***
## year_2013        -2.893e-02  1.765e-01  -0.164    0.870
## year_2014         1.534e-01  1.660e-01   0.924    0.356
## year_2015         2.989e-02  1.596e-01   0.187    0.851
## year_2016        -4.282e-02  1.599e-01  -0.268    0.789
## year_2017        -4.455e-02  1.774e-01  -0.251    0.802
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.384 on 1374 degrees of freedom
```

```
## Multiple R-squared:   0.73,  Adjusted R-squared:  0.7275
## F-statistic: 285.8 on 13 and 1374 DF,  p-value: < 2.2e-16
```

Interpretation (year_2010): Our model suggests that Markup is 0.175 units larger in 2010 compared to 2009. However, the effect is not significantly different from zero.

The time trend would be misleading since we have only one exceptional year driving the time trend effect (2012).

We might also include all time dummies and drop the intercept

```
OLSni =
lm(Markup~RevGR+eqshare+FCR+age+TotalassetsthEUR+year_2009+year_2010+ye
ar_2011

+year_2012+year_2013+year_2014+year_2015+year_2016+year_2017-
1,dat=mapo)
summary(OLSni)

##
## Call:
## lm(formula = Markup ~ RevGR + eqshare + FCR + age + TotalassetsthEUR
+
##     year_2009 + year_2010 + year_2011 + year_2012 + year_2013 +
##     year_2014 + year_2015 + year_2016 + year_2017 - 1, data = mapo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9608 -0.6057 -0.1581  0.3522 22.4378
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## RevGR              1.159e-03  1.665e-03   0.696    0.487
## eqshare            2.725e-03  1.928e-03   1.414    0.158
## FCR                2.647e-02  3.664e-03   7.224 8.32e-13 ***
## age                1.912e-03  1.556e-03   1.229    0.219
## TotalassetsthEUR -1.205e-07  5.623e-07  -0.214    0.830
## year_2009          2.555e+00  1.696e-01  15.068  < 2e-16 ***
## year_2010          2.730e+00  1.537e-01  17.758  < 2e-16 ***
## year_2011          2.525e+00  1.588e-01  15.902  < 2e-16 ***
## year_2012         -5.543e+00  1.721e-01 -32.206  < 2e-16 ***
## year_2013          2.526e+00  1.663e-01  15.189  < 2e-16 ***
## year_2014          2.708e+00  1.467e-01  18.460  < 2e-16 ***
## year_2015          2.585e+00  1.570e-01  16.466  < 2e-16 ***
## year_2016          2.512e+00  1.532e-01  16.395  < 2e-16 ***
## year_2017          2.510e+00  1.644e-01  15.272  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.384 on 1374 degrees of freedom
```

```
## Multiple R-squared:  0.8628, Adjusted R-squared:  0.8614
## F-statistic: 617.2 on 14 and 1374 DF,  p-value: < 2.2e-16
```

If we did not drop the intercept when including all dummies, R automatically drops a random dummy from the regression to avoid perfect multicollinearity.

```
OLSfai =
lm(Markup~RevGR+eqshare+FCR+age+TotalassetsthEUR+year_2009+year_2010+ye
ar_2011

+year_2012+year_2013+year_2014+year_2015+year_2016+year_2017,dat=mapo)
summary(OLSfai)

##
## Call:
## lm(formula = Markup ~ RevGR + eqshare + FCR + age + TotalassetsthEUR
+
##     year_2009 + year_2010 + year_2011 + year_2012 + year_2013 +
##     year_2014 + year_2015 + year_2016 + year_2017, data = mapo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9608 -0.6057 -0.1581  0.3522 22.4378
##
## Coefficients: (1 not defined because of singularities)
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.510e+00  1.644e-01  15.272  < 2e-16 ***
## RevGR             1.159e-03  1.665e-03   0.696    0.487
## eqshare           2.725e-03  1.928e-03   1.414    0.158
## FCR               2.647e-02  3.664e-03   7.224 8.32e-13 ***
## age               1.912e-03  1.556e-03   1.229    0.219
## TotalassetsthEUR -1.205e-07  5.623e-07  -0.214    0.830
## year_2009         4.455e-02  1.774e-01   0.251    0.802
## year_2010         2.194e-01  1.623e-01   1.352    0.177
## year_2011         1.451e-02  1.646e-01   0.088    0.930
## year_2012        -8.053e+00  1.784e-01 -45.137  < 2e-16 ***
## year_2013         1.562e-02  1.726e-01   0.090    0.928
## year_2014         1.979e-01  1.606e-01   1.232    0.218
## year_2015         7.443e-02  1.610e-01   0.462    0.644
## year_2016         1.730e-03  1.606e-01   0.011    0.991
## year_2017                NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.384 on 1374 degrees of freedom
## Multiple R-squared:   0.73,  Adjusted R-squared:  0.7275
## F-statistic: 285.8 on 13 and 1374 DF,  p-value: < 2.2e-16
```

As a last example, let us include a slope dummy and intercept dummy (for 2012)

```r
OLSsld =
lm(Markup~RevGR+eqshare+FCR+age+TotalassetsthEUR+year_2010+year_2011

+year_2012+year_2013+year_2014+year_2015+year_2016+year_2017
            +year_2012:TotalassetsthEUR,dat=mapo)
summary(OLSsld)
```

```
##
## Call:
## lm(formula = Markup ~ RevGR + eqshare + FCR + age + TotalassetsthEUR
+
##     year_2010 + year_2011 + year_2012 + year_2013 + year_2014 +
##     year_2015 + year_2016 + year_2017 + year_2012:TotalassetsthEUR,
##     data = mapo)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.9382 -0.6062 -0.1561  0.3524 22.4392
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  2.555e+00  1.696e-01  15.063  < 2e-16
***
## RevGR                        1.160e-03  1.666e-03   0.696    0.486
## eqshare                      2.726e-03  1.928e-03   1.414    0.158
## FCR                          2.645e-02  3.668e-03   7.210 9.19e-13
***
## age                          1.909e-03  1.557e-03   1.226    0.220
## TotalassetsthEUR            -9.493e-08  5.770e-07  -0.165    0.869
## year_2010                    1.748e-01  1.647e-01   1.062    0.289
## year_2011                   -3.013e-02  1.694e-01  -0.178    0.859
## year_2012                   -8.088e+00  1.859e-01 -43.495  < 2e-16
***
## year_2013                   -2.939e-02  1.765e-01  -0.166    0.868
## year_2014                    1.528e-01  1.661e-01   0.920    0.358
## year_2015                    2.966e-02  1.597e-01   0.186    0.853
## year_2016                   -4.320e-02  1.599e-01  -0.270    0.787
## year_2017                   -4.521e-02  1.775e-01  -0.255    0.799
## TotalassetsthEUR:year_2012 -5.075e-07  2.558e-06  -0.198    0.843
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.385 on 1373 degrees of freedom
## Multiple R-squared:   0.73,  Adjusted R-squared:  0.7273
## F-statistic: 265.2 on 14 and 1373 DF,  p-value: < 2.2e-16
```