

# Multiple Linear Regression

Multiple linear regression is an extension of simple linear regression used to predict an outcome variable (y) on the basis of multiple distinct predictor variables (x).

With three predictor variables (x), the prediction of y is expressed by the following equation:

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3$$

The “b” values are called the regression weights or beta coefficients. They measure the association between the predictor variable and the outcome.

```
> summary(model)$coefficient
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.526667243	0.374289884	9.4222884	1.267295e-17
youtube	0.045764645	0.001394897	32.8086244	1.509960e-81
facebook	0.188530017	0.008611234	21.8934961	1.505339e-54
newspaper	-0.001037493	0.005871010	-0.1767146	8.599151e-01

```
> |
```

```
##
```

- It is noted changing in you tube and face book advertising budget are significantly associated to changes in sales while changes in newspaper budget is not significantly associated with sales.
- For a fixed amount of you tube and newspaper advertising budget, spending an additional 1000 dollars on facebook advertising leads to an increase in sales by approximately  $0.1885 \times 1000 = 189$  sale units, on average.
- For you tube coefficient suggests that for every 1000 dollars increase in you tube advertising budget, holding all other predictors constant, we can expect an increase of  $0.045 \times 1000 = 45$  sales units, on average.

# Model building

- Newspaper is not significant in the multiple regression model.
- This means that, for a fixed amount of you tube and newspaper advertising budget, changes in the newspaper advertising budget will not significantly affect sales units.
- As the newspaper variable is not significant, it is possible to remove it from the model:
- $\text{sales} = 3.5 + 0.045 \cdot \text{youtube} + 0.187 \cdot \text{facebook}$ .
- With youtube and facebook predictor variables, the adjusted  $R^2 = 0.89$ , meaning that “89% of the variance in the measure of sales can be predicted by youtube and facebook advertising budgets.

- ```
model <- lm(sales ~ youtube + facebook, data = marketing)
summary(model)
```

```
##
## Call:
## lm(formula = sales ~ youtube + facebook, data = marketing)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.557  -1.050   0.291   1.405   3.399
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.50532    0.35339   9.92  <2e-16 ***
## youtube       0.04575    0.00139  32.91  <2e-16 ***
## facebook      0.18799    0.00804  23.38  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.02 on 197 degrees of freedom
## Multiple R-squared:  0.897, Adjusted R-squared:  0.896
## F-statistic: 860 on 2 and 197 DF, p-value: <2e-16
```

# To compute multiple regression

- To compute multiple regression using all of the predictors in the data set
- `model <- lm(sales ~., data = marketing)`
- If we want to perform regression using all the variables except one
- `model <- lm(sales ~. -newspaper, data = marketing)`