3.6 Interactions

3.6.1 Definition

Interactions occur when any given predictor variable has a different effect on the response variable depending on the value of a different predictor.

3.6.2 Types of Interactions

Resulting from the different types of predictor variables (continuous or categorical), the following different types of interactions might be considered

- continuous by continuous
- continuous by categorical
- categorical by categorical

For reasons of simplicity, we start with the second category

3.6.3 Continuous by Categorical

The continuous by categorical interaction occurs when the linear model contains two predictors, one is a regression variable and the second predictor is a factor with categorical levels. In our example dataset with body weight of beef cattle animals this would be the case when `Breast Circumference` and `Breed` are both included into the same model. The example dataset that is used to show this type of interactions is shown below.

Table 3.9: Body Weight, Breast Circumference and Breed of Beef Cattle Animals

Animal	Breast Circumference	Body Weight	Breed
1	176	471	Angus
2	177	463	Angus
4	179	470	Angus
7	181	518	Limousin
8	182	511	Limousin
9	183	510	Limousin
10	184	541	Limousin
3	178	481	Simmental
5	179	496	Simmental
6	180	491	Simmental

The regression of `Body Weight` on `Breast Circumference` is extended by the categorical variable `Breed`. This leads to the following results of the summary-output of `lm()`.

```
fm_bw_br_bc <- `Body Weight` ~ Breed + `Breast Circumference`
lm__bw_br_bc <- lm(formula = fm_bw_br_bc, data = tbl_flem_bw_bc_br)
(smry_lm_bw_br_bc <- summary(lm__bw_br_bc))
```

```

Call:
lm(formula = fm_bw_br_bc, data = tbl_flem_bw_bc_br)

Residuals:
    Min      1Q  Median      3Q     Max 
-11.929  -4.464  -2.952   5.946  15.214 

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)            -216.000    522.272  -0.414    0.694
BreedLimousin            32.071     17.045   1.882    0.109
BreedSimmental           14.905      9.568   1.558    0.170
`Breast Circumference`    3.857      2.945   1.310    0.238

Residual standard error: 10.06 on 6 degrees of freedom
Multiple R-squared:  0.8909,	Adjusted R-squared:  0.8363 
F-statistic: 16.33 on 3 and 6 DF,  p-value: 0.002723
```

Here, it is important to note that the order of the terms in the linear model formula does not matter

If we have reasons to assume that the regression of `Body Weight` on `Breast Circumference` is different for the different breeds, we can include that fact into our model by a so-called **interaction term** `BreastCircumference`:`Breed`. The model formulation and the summary of the results as obtained by the function `lm()` is shown below.

```
fm_bw_bc_br_int <- `Body Weight` ~ `Breast Circumference` +
  Breed + `Breast Circumference`:Breed
lm__bw_bc_br_int <- lm(formula = fm_bw_bc_br_int, data = tbl_flem_bw_bc_br)
(smry_lm_bw_bc_br_int <- summary(lm__bw_bc_br_int))
```

```

Call:
```

```
lm(formula = fm_bw_bc_br_int, data = tbl_flem_bw_bc_br)

Residuals:
      1       2       3       4       5       6       7       8       9      10 
  3.286  -4.929   1.643   8.200  -5.600 -13.400  10.800  -3.333   6.667  -3.333 

Coefficients:
                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                             430.0000   917.1235   0.469    0.664
`Breast Circumference`                    0.2143     5.1716   0.041    0.969
BreedLimousin                         -1151.0000  1293.2741  -0.890    0.424
BreedSimmental                         -835.6667  1685.4451  -0.496    0.646
`Breast Circumference`:BreedLimousin      6.5857     7.1908   0.916    0.412
`Breast Circumference`:BreedSimmental     4.7857     9.4420   0.507    0.639

Residual standard error: 11.17 on 4 degrees of freedom
Multiple R-squared:  0.9103,	Adjusted R-squared:  0.7981 
F-statistic: 8.115 on 5 and 4 DF,  p-value: 0.03212
```

The model matrix for the interaction model is given by

```
model.matrix(lm__bw_bc_br_int)
```

```
   (Intercept) `Breast Circumference` BreedLimousin BreedSimmental
1            1                    176             0              0
2            1                    177             0              0
3            1                    179             0              0
4            1                    181             1              0
5            1                    182             1              0
6            1                    183             1              0
7            1                    184             1              0
8            1                    178             0              1
9            1                    179             0              1
10           1                    180             0              1
   `Breast Circumference`:BreedLimousin `Breast Circumference`:BreedSimmental
1                                     0                                     0
2                                     0                                     0
3                                     0                                     0
4                                   181                                     0
5                                   182                                     0
6                                   183                                     0
7                                   184                                     0
```

```
8                                      0                                    178
9                                      0                                    179
10                                     0                                    180
attr(,"assign")
[1] 0 1 2 2 3 3
attr(,"contrasts")
attr(,"contrasts")$Breed
[1] "contr.treatment"
```

The normal equations for the model with interactions can be established and solved as shown below.

```
mat_X <- model.matrix(lm__bw_bc_br_int)
attr(mat_X, "assign") <- NULL
attr(mat_X, "contrasts") <- NULL
mat_xtx <- crossprod(mat_X)
mat_xty <- crossprod(mat_X, tbl_flem_bw_bc_br$`Body Weight`)
mat_b_sol <- solve(mat_xtx, mat_xty)
mat_b_sol
```

```
                                            [,1]
(Intercept)                          430.0000000
`Breast Circumference`                 0.2142857
BreedLimousin                      -1151.0000000
BreedSimmental                      -835.6666667
`Breast Circumference`:BreedLimousin   6.5857143
`Breast Circumference`:BreedSimmental  4.7857143
```

To better understand the meaning of the above shown results, we are having a closer look at the structure of the model with an interaction term. From a modelling point-of-view, the interaction can also be seen as a dependency of the regression coefficient b_1 on the breed. Hence, the original restriced model without interaction which is given by

$$y_i = b_0 + b_1 \times BC_i + b_2 \times BrLi_i + b_3 \times BrSi_i + e_i$$

can be extended to include the linear relationship of b_1 with `Breed`

$$b_1 = a + b_4 \times BrLi + b_5 \times BrSi$$

Inserting that into the model gives

$$y_i = b_0 + (a + b_4 \times BrLi + b_5 \times BrSi) \times BC_i + b_2 \times BrLi_i + b_3 \times BrSi_i + e_i$$

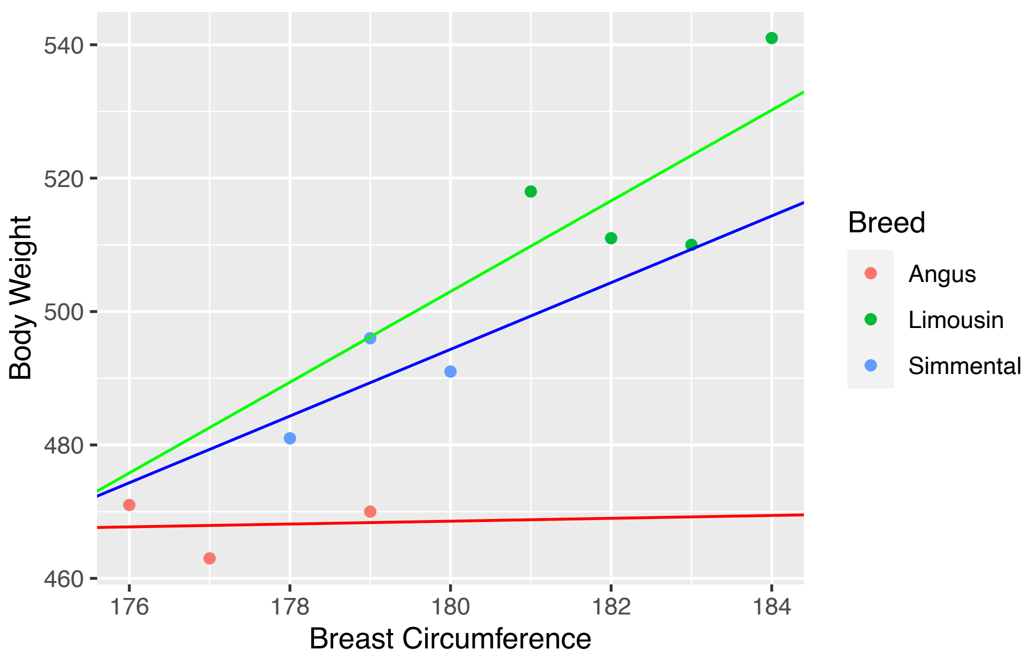
which can be simplified to

$$y_i = b_0 + a \times BC_i + b_2 \times BrLi_i + b_3 \times BrSi_i + b_4 \times BrLi \times BC_i + b_5 \times BrSi \times BC_i + e_i$$

An increase in 1 cm in Breast Circumference for an Angus animal leads to 0.21 kg inrease in body weight. For a Limousin animal this increase is $0.21 + 6.59 = 6.8$. For a Simmental animal, the increase is $0.21 + 4.79 = 5$

The mean Body Weight of Angus animals with $BC = 0$ is given by the intercept and correpsonds to 430.

The following plot should explain the relationship, between the different slope estimates. The regression lines are estimated for the different breeds separately. The different slopes of the breed-specific regression lines indicates that there are interactions between the predictors `Breed` and `Breast Circumference`.



3.6.4 Continuous by Continuous

The effect of a continuous by continuous interaction can be explained similarly to what was shown above in the continuous by categorical case. We have a model with a multiple linear regression such as `Body Weight` on `Breast Circumference` and `Height`. Such a multiple linear regression model without interactions can be written as

$$y_0 = b_0 + b_1 \times BC_i + b_2 \times HE_i + e_i \tag{3.17}$$

Extending the above model to the case with interactions can be done by assuming that the regression coefficient b_1 for `Breast Circumference` has a linear dependence to `Height`. Then b_1 can be written as

$$b_1 = b_3 + b_4 \times HE_i$$

Inserting this into the model without interactions leads to

$$y_0 = b_0 + (b_3 + b_4 \times HE_i) \times BC_i + b_2 \times HE_i + e_i$$

After simplifications, we get

$$y_0 = b_0 + b_2 \times HE_i + b_3 \times BC_i + b_4 \times HE_i \times BC_i + e_i$$

Hence the categorical by categorical interaction is a regression coefficient that shows the effect of the product of multiple linear regression predictors.

3.6.5 Categorical by Categorical

The categorical by categorical interaction occurs when the effects of a given factor such as `Breed` depends the level of a different factor such as `Sex`.

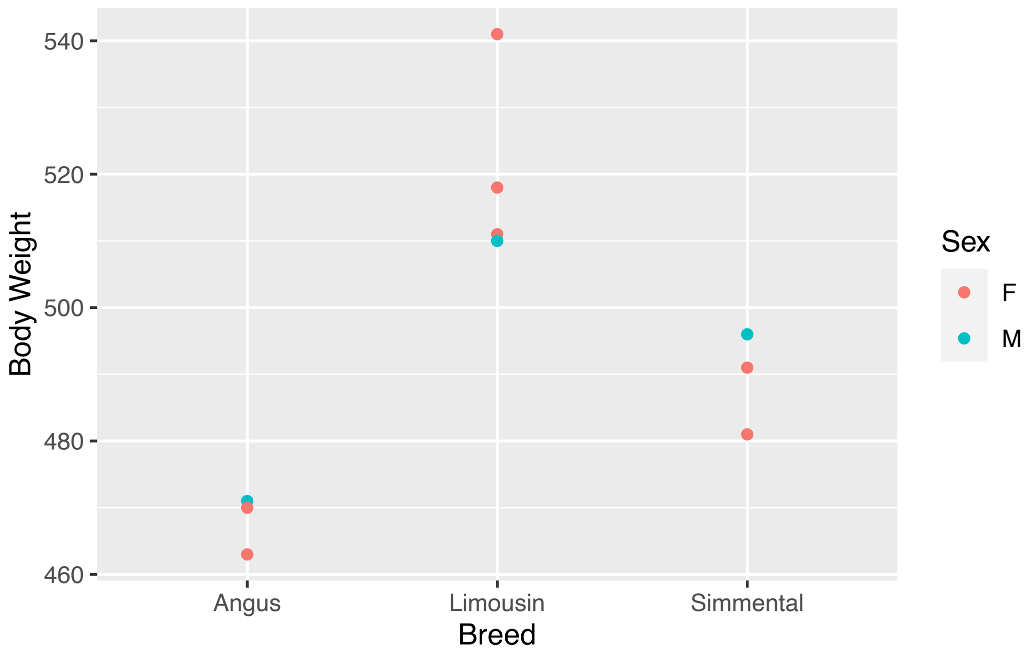
Let us assume that we have a dataset with `Body Weight` as response and `Breed` and `Sex` as predictor variables. Such an example dataset is shown in Table Table 3.10.

Table 3.10: Body Weight, Breed and Sex for different Beef Cattle Animals

Animal	Body Weight	Breed	Sex
1	471	Angus	M
2	463	Angus	F
4	470	Angus	F
7	518	Limousin	F
8	511	Limousin	F
9	510	Limousin	M
10	541	Limousin	F

Animal	Body Weight	Breed	Sex
3	481	Simmental	F
5	496	Simmental	M
6	491	Simmental	F

The most instructive way to see the effect of an interaction between two categorical predictors is to visualise the data using a plot where one predictor is shown on the x-axis and the second predictor is used as grouping variable



In the above plot, we can see that for the breeds `Angus` and `Simmental`, male animals are heavier than female animals. In contrast, for the breed `Limousin` female animals are heavier than male animals. This gives a strong hint to an interaction between `Sex` and `Breed`.