6.4 Interaction Terms

In a few words: In the different types of regression models we saw in Chapter 5, the effects of the predictors X_i are added together to explain the outcome Y. This is also the case in the following logistic regression model where Y is "laryngeal cancer" (yes[1]/no[0]), X_1 is "alcohol consumption" (yes[1]/no[0]), and X_2 is "tobacco consumption" (yes[1]/no[0]): $Log\left[\dfrac{prob(Y=1)}{1-prob(Y=1)}\right]$ $= a_0 + a_1 \times X_1 + a_2 \times X_2$. This relationship has a non-trivial implication: The effect of alcohol on the risk of presenting a laryngeal cancer is the same whether or not a subject smokes. Unfortunately, this does not correspond to what is commonly observed (Flanders and Rothman 1982): "The exposure to both factors increases the risk about 50% more than the increase predicted if the effects of tobacco and alcohol were simply additive." In such a situation, it is said that there is an "interaction" between alcohol and tobacco and that this interaction is positive because the exposure to both factors increases the risk.

From a statistical point of view, a simple way to capture certain elements of interaction between X_1 and X_2 consists of adding the product $X_1 \times X_2$ to the model. In our previous example, this now gives the following:

$$Log\left[\frac{prob(Y=1)}{1-prob(Y=1)}\right] = a_0 + a_1 \times X_1 + a_2 \times X_2 + a_3 \times X_1 \times X_2,$$

so that in smokers ($X_2 = 1$), we have

$$Log\left[\frac{prob(Y=1)}{1-prob(Y=1)}\right] = a_0 + a_2 \times 1 + a_1 \times X_1 + a_3 \times X_1 \times 1 = (a_0 + a_2) + (a_1 + a_3) \times X_1$$

and in non-smokers ($X_2 = 0$), we have

$$Log\left[\frac{prob(Y=1)}{1-prob(Y=1)}\right] = a_0 + a_1 \times X_1.$$

There is now a different level of association between the risk of presenting laryngeal cancer and alcohol consumption, depending on whether or not a subject smokes. In practice, following this approach, the test of the hypothesis $a_3 = 0$ will formally answer the following question: Is there a statistically significant interaction between X_1 and X_2?

Two problems are commonly encountered in the management of interaction terms in statistical models. The first concerns the identification of pairs of variables for which an interaction should be considered. It is often

a disconcerting and unmanageable task to look carefully for each possible case. Here again, a review of the literature on the particular topic addressed by the model will help find the main potential interactions. An interesting approach proposed by Cox and Wermuth (1996) is presented in the practical part of this section. It helps determine globally whether or not a set of outcomes and predictors raises the question of interaction terms.

The second problem concerns the interpretation of the so-called main effects X_1 (alcohol consumption) and X_2 (tobacco consumption) in the presence of the interaction $X_1 \times X_2$. This point is regularly discussed in the literature (Rothman and Greenland 1998; Fox 2002) and there is no real consensus. Here are some elements of this debate:

- In our example, the test of $a_1 = 0$ involves the test of the effect of alcohol consumption on the risk of presenting a laryngeal cancer in a population of non-smokers (because in this situation $X_2 = 0$ so that the interaction term disappears). This interpretation is possible because $X_2 = 0$ has a clear meaning here. This is not always the case, for example if the variable X_1 was the "age": age = 0 is birth, which is not necessarily a relevant reference. In this latter situation, it is sometimes suggested (Rothman and Greenland 1998) that "age" should be re-centered, for example by using the relation "age2 = age − 18" or "age2 = age − 65", if the age of 18 or 65 is considered a milestone disease (depending on the question under study).

  AU: Stet OR "milestone of disease"?

- The previous option has a drawback: The test of $a_1 = 0$ does not involve the effect of alcohol consumption "in general" but only in a particular population. This is correct from a statistical point of view, but it can be frustrating from a public health perspective. An alternative sometimes proposed is to center X_1 and X_2 by their mean; more precisely, X_1 is replaced by $X'_1 = X_1 − \text{mean}(X_1)$ and X_2 is replaced by $X'_2 = X_2 − \text{mean}(X_2)$. In this situation, the test of $a_1 = 0$ concerns the effect of alcohol consumption in a virtual population of subjects with a prevalence of tobacco consumption similar to the prevalence observed in the study (Koopman 2008). The interpretation is perhaps more intuitive, but this approach implies a manipulation of data and it is difficult to generalize to non-binary categorical variables.

- A third possibility (which is applicable to categorical variables) is to recode variables X_1 and X_2 as (1,−1) instead of (1,0) or, more generally and following the R taxonomy, to use a "contr.sum" option instead of the traditional "contr.treatment" one. Here, the test of $a_1 = 0$ involves the effect of alcohol consumption in a virtual population of subjects with a prevalence of tobacco consumption of 50%.

- Finally, some authors are more straightforward and suggest estimating the main effects using a model where the interaction term has been simply discarded (Fox 2002). This makes sense, even if it raises

a theoretical problem: If the interaction does exist and if the interaction term is not in the model, then it is in the residual. But in this case, the residual is no longer "pure noise" (if the residual contains the interaction term $X_1 \times X_2$, then it is not independent of X_1 and X_2).

In Practice: In the previous section we designed a series of models that explained the probability of being at high risk for suicide attempt. If, for instance, we consider the model obtained using the stepwise selection procedure, ten explanatory variables were selected. Among them was the variable named "direct", which stands for a low self-directedness (yes[1]/no[0]; low self-directedness is defined by poor impulse control or a weak ego (Cloninger 2000)). The variable "direct" was positively and significantly associated with the outcome, the odds-ratio being equal to exp(1.18❽) = 3.3). Another variable was named "f.type.center", which stands for the type of prison: high-security unit (type 1), centre for longer sentences or for prisoners with good scope for rehabilitation (type 2), or a prison for remand prisoners or short sentences (type 3). This variable was not significantly associated with the outcome. It could be hypothesised that low self-directedness is particularly deleterious in high-security prisons as regards risk of suicide attempt, so that this could correspond formally to an interaction between "direct" and "f.type. center" in explaining "suicide.hr".

Let us now see how to estimate and test this interaction in the corresponding logistic regression model:

```
                    ❶
> str(mhp.mod$direct)
 num [1:799] 0 0 0 0 0 0 0 0 0 1 0 ...
                           ❷
> contrasts(mhp.mod$f.type.center)
    2   3
1   0   0❸
2   1   0
3   0   1
> mod <- glm(suicide.hr ~ grav.cons + suicide.past + dep.cons
  + past.prison + sczdys.cons + death + against.people
  + subst.cons + direct + f.type.center
              ❹
  + direct:f.type.center, data = mhp.mod, family = "binomial")
> summary(mod)

Call:
glm(formula = suicide.hr ~ grav.cons + suicide.past + dep.cons
  + past.prison + sczdys.cons + death + against.people
  + subst.cons + direct + f.type.center + direct:f.type.center,
  family = "binomial", data = mhp.mod)
```

```
Deviance Residuals:
Min        1Q        Median      3Q         Max
-2.03462   -0.43702  -0.19137   -0.07424   3.16051

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)           -6.0626     0.7330    -8.271  < 2e-16 ***
grav.cons              0.6692     0.1207     5.545 2.94e-08 ***
suicide.past           1.7729     0.2791     6.352 2.13e-10 ***
dep.cons               1.4319     0.2893     4.950 7.43e-07 ***
past.prison            0.7321     0.2867     2.554  0.0107 *
sczdys.cons            1.3722     0.8377     1.638  0.1014
death                  0.3008     0.2868     1.049  0.2943
against.people        -0.2359     0.2984    -0.791  0.4292
subst.cons            -0.2405     0.3100    -0.776  0.4380
direct                -1.4466❺    1.0311    -1.403  0.1606
f.type.center2        -0.5643     0.4960    -1.138  0.2553
f.type.center3        -0.3066     0.4624    -0.663  0.5073
direct:f.type.center2  2.8487❻    1.1701     2.435  0.0149 *
direct:f.type.center3  2.7083❼    1.1012     2.459  0.0139 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 659.51 on 663 degrees of freedom
Residual deviance: 370.15 on 650 degrees of freedom
 (135 observations deleted due to missingness)
AIC: 398.15

Number of Fisher Scoring iterations: 6
    ❽
> drop1(mod, .~., test = "Chisq")
Single term deletions

Model:
suicide.hr ~ grav.cons + suicide.past + dep.cons + past.prison
  + sczdys.cons + death + against.people + subst.cons + direct
  + f.type.center + direct:f.type.center
                   Df  Deviance   AIC    LRT   Pr(Chi)
<none>                  370.15  398.15
grav.cons          1    405.54  431.54 35.38 2.706e-09 ***
suicide.past       1    413.24  439.24 43.09 5.220e-11 ***
dep.cons           1    396.28  422.28 26.13 3.192e-07 ***
past.prison        1    376.80  402.80  6.65 0.00992    **
sczdys.cons        1    373.15  399.15  3.00 0.08343    .
```

```
death                   1   371.25  397.25  1.10 0.29523
against.people          1   370.78  396.78  0.63 0.42900
subst.cons              1   370.76  396.76  0.61 0.43528
direct                  1   371.09  397.09  0.94 0.33183   ❾
f.type.center           2   371.49  395.49  1.34 0.51282
direct:f.type.center    2   378.03  402.03  7.88 0.01947   *❿
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It is first necessary to verify the coding of the two variables present in the interaction term; "direct"❶ and "f.type.center"❷ are coded here as (1,0). Because "f.type.center" is a categorical variable with three levels, it is recoded with two binary variables❷. In the function glm(), the interaction term "direct × f.type.center" is represented by direct:f.type.center❹. The global effect of each variable is obtained from the instruction drop1()❽. The interaction "direct × f.type.center" is statistically significant❿; the main effect "direct" is no longer significant❾ but this should be considered very cautiously (and perhaps not taken into consideration at all) because this is a main effect in the presence of an interaction term that contains it. Always about variable "direct", how can the coefficient ❺ be interpreted?. As explained above, the associated odds-ratio equal to exp(–1.45❺) = 0.24 is related to the sub-population of prisoners for which f.type.center2 = 0 and f.type.center3 = 0. From ❸ this corresponds to type 1 prisons 1 (high-security units): a low level of self-directedness is associated with a risk of suicide attempt divided by 4.17 (1/4.17 = 0.24) in high-security units (we had hypothesised the reverse). However, this association is not statistically significant (p = 0.16). For prisoners in type 2 prisons (centres for longer sentences or for prisoners with good scope for rehabilitation), from ❸ we note that f.type.center2 = 1 and f.type.center3 = 0 so that the odds-ratio is now exp(–1.45❺ + 2.85❻) = 4.1. Here, a low level of self-directedness is a risk factor for suicide attempt. For prisoners in type 3 prisons, we have exp(–1.45❺ + 2.71❻) = 3.5. To test whether or not these last two odds-ratios are statistically different from 1, we can use the function estimable() in the package "gmodels" (see Section 6.2):

```
> library(gmodels)
> estimable(mod, c(0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0),
  conf.int = 0.95)
                               Estimate   Std. Error X^2 value DF
(0 0 0 0 0 0 0 0 0 1 0 0 1 0)  1.402069   0.5368913  6.819711  1
                               Pr(>|X^2|)  Lower.CI   Upper.CI
(0 0 0 0 0 0 0 0 0 1 0 0 1 0) ❶0.009015716 0.3368918  2.467246
> estimable(mod, c(0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1),
  conf.int = 0.95)
```

```
                                 Estimate   Std. Error X^2 value DF
(0 0 0 0 0 0 0 0 0 1 0 0 0 1)   1.261613   0.3877241 10.58782  1
                               Pr(>|X^2|)   Lower.CI  Upper.CI
(0 0 0 0 0 0 0 0 0 1 0 0 0 1) ❷0.001138351 0.4923792 2.030846
```

Both p-values are lower than 0.05 in ❶ and ❷.

Now if we use a (1,–1) coding for the variables "direct" and "f.type.prison", we obtain

```
                              ❶
> mhp.mod$direct2 <- ifelse(mhp.mod$direct == 1, 1, -1)
      ❷
> table(mhp.mod$direct, mhp.mod$direct2, useNA = "ifany")


        -1    1   <NA>
0      567    0     0
1        0  147     0
<NA>     0    0    85


                                           ❸
> contrasts(mhp.mod$f.type.center) <- contr.sum
> contrasts(mhp.mod$f.type.center)
      [, 1]  [, 2]
1        1      0
2        0      1
3       -1     -1
> mod <- glm(suicide.hr ~ grav.cons + suicide.past
  + dep.cons + past.prison + sczdys.cons + death
  + against.people + subst.cons + direct2 + f.type.center
  + direct2:f.type.center, data = mhp.mod, family = "binomial")
> summary(mod)

Call:
glm(formula = suicide.hr ~ grav.cons + suicide.past + dep.cons
  + past.prison + sczdys.cons + death + against.people
  + subst.cons + direct2 + f.type.center + direct2:f.type.center,
  family = "binomial", data = mhp.mod)

Deviance Residuals:
  Min        1Q      Median       3Q        Max
-2.03462  -0.43702  -0.19137  -0.07424   3.16051

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)         -6.1500     0.6633   -9.272  < 2e-16 ***
grav.cons            0.6692     0.1207    5.545 2.94e-08 ***
suicide.past         1.7729     0.2791    6.352 2.13e-10 ***
```

```
dep.cons                 1.4319    0.2893    4.950  7.43e-07 ***
past.prison              0.7321    0.2867    2.554    0.0107 *
sczdys.cons              1.3722    0.8377    1.638    0.1014
death                    0.3008    0.2868    1.049    0.2943
against.people          -0.2359    0.2984   -0.791    0.4292
subst.cons              -0.2405    0.3100   -0.776    0.4380
direct2                  0.2028❹   0.2031    0.999    0.3179
f.type.center1          -0.6359    0.3681   -1.728    0.0840 .
f.type.center2           0.2242    0.2617    0.857    0.3916
direct2:f.type.center1  -0.9262    0.3623   -2.557    0.0106 *
direct2:f.type.center2   0.4982    0.2584    1.928    0.0538 .
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 659.51 on 663 degrees of freedom
Residual deviance: 370.15 on 650 degrees of freedom
 (135 observations deleted due to missingness)
AIC: 398.15


Number of Fisher Scoring iterations: 6

> drop1(mod,.~.,test="Chisq")
Single term deletions

Model:
suicide.hr ~ grav.cons + suicide.past + dep.cons + past.prison
  + sczdys.cons + death + against.people + subst.cons
  + direct2 + f.type.center + direct2:f.type.center
                      Df  Deviance    AIC    LRT   Pr(Chi)
<none>                    370.15   398.15
grav.cons             1  405.54   431.54  35.38  2.706e-09 ***
suicide.past          1  413.24   439.24  43.09  5.220e-11 ***
dep.cons              1  396.28   422.28  26.13  3.192e-07 ***
past.prison           1  376.80   402.80   6.65    0.00992 **
sczdys.cons           1  373.15   399.15   3.00    0.08343 .
death                 1  371.25   397.25   1.10    0.29523
against.people        1  370.78   396.78   0.63    0.42900
subst.cons            1  370.76   396.76   0.61    0.43528
direct2               1  371.09   397.09   0.94    0.33183  ❺
f.type.center         2  374.14   398.14   3.99    0.13605  ❻
direct2:f.type.center 2  378.03   402.03   7.88    0.01947 *❼
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The variable "direct" is recoded in ❶, and the modification is verified in ❷. The variable "f.type.center" is recoded in ❸ with the "contr.sum" option of the `contrasts()` function (which could also have been used with "direct" if it had been defined as a factor instead of a binary numerical variable). The `drop1()` function leads to similar results for the interaction term ❼ with the two codings (and this is expected), but gives different estimates of p-values for the main effect ❻. The coefficient of "direct" is equal to 0.2028❹. It can be interpreted as follows: Because the coding is (1,–1), the odds-ratio associated with this variable is exp(2 × 0.2028) = 1.5. This odds-ratio involves the effect of low self-directedness on being at high risk of suicide attempt, in a population where the three types of prison are assumed to be balanced, and given that the other predictor variables in the model are held constant.

Now it is time to have a look at the procedure proposed by Cox and Wermuth (1996). This procedure gives an idea of the overall importance of interaction terms in a given model. First, all interaction terms are estimated and tested, one at a time. Second, the corresponding p-values are plotted in a diagram, and a visual assessment can help determine if there is indeed a substantial amount of interaction terms that should be retained. It would be tedious to perform all these computations using a classical statistical package. With R, it takes only a few lines:

```
            ❶
> mod <- glm(suicide.hr ~ grav.cons + suicide.past + dep.cons
  + direct + past.prison + sczdys.cons + death + against.people
  + subst.cons + f.type.center, data = mhp.mod,
  family = "binomial")
                          ❷              ❸            ❹
> mod.inter <- data.frame(add1(mod, ~.^2, test = "Chisq"))
                ❺                        ❻
> isort <- order(mod.inter["Pr.Chi."])
                      ❼
> mod.inter[isort, ]
                           Df Deviance    AIC       LRT      Pr.Chi.
past.prison:against.people 1 370.8583 396.8583 7.171484133 0.007407147❽
suicide.past:dep.cons      1 371.6522 397.6522 6.377570255 0.011557154
direct:f.type.center       2 370.1516 398.1516 7.878152389 0.019466189
dep.cons:sczdys.cons       1 373.7124 399.7124 4.317342548 0.037725795
suicide.past:direct        1 374.5811 400.5811 3.448631713 0.063304155
grav.cons:against.people   1 375.2116 401.2116 2.818140682 0.093204312
dep.cons:against.people    1 375.3861 401.3861 2.643630152 0.103965739
grav.cons:dep.cons         1 375.5737 401.5737 2.456018505 0.117075207
sczdys.cons:f.type.center  2 373.7939 401.7939 4.235870649 0.120279711
against.people:subst.cons  1 375.9531 401.9531 2.076679837 0.149565193
past.prison:f.type.center  2 374.2442 402.2442 3.785512614 0.150655982
.... [TRUNCATED]
```
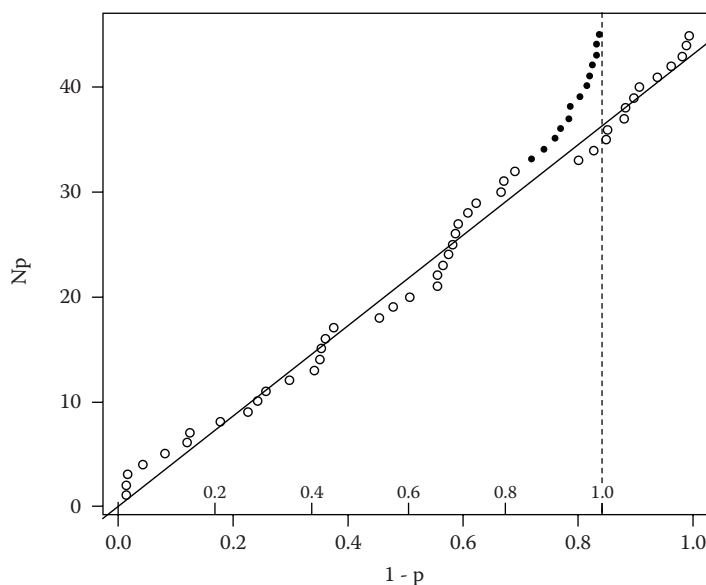
FIGURE 6.3
Graphical representation of the p-values of the forty-five possible interaction terms in a logistic regression model. Under the assumption that all interaction terms are null, the empty dots should be on a straight line (each dot corresponds to an interaction). This is observed in the present situation. The solid dots correspond to an imaginary situation where there seems to be surprisingly small p-values for some interaction terms.

The basic model without any interaction is in ❶. The function add1() ❸ is then called to add all interaction terms as specified in ❹, one at a time. The corresponding results are stored in a data frame❷. The index of the line with the smallest p-value❻, the next smallest, etc. is obtained using the function order(). The sorted results are presented in ❼; the interaction term with the smallest p-value is in ❽.

These p-values can be plotted (Figure 6.3):

```
                        ❶
> x <- 1-na.omit(mod.inter[isort, "Pr.Chi."])
            ❷
> y <- length(x):1
> plot(x, y, xlab = "1 - p", ylab = "Np")
            ❸                        ❹
>     line <- lm(y[length(x):1] ~ -1 + x[length(x):1])
      ❺
> abline(line)
```

In this plot (Figure 6.3), the ordered p-values❶ obtained previously are represented on the x-axis, and the numbers of the interaction terms❷ are

represented on the y-axis, ranked from the largest p-value to the smallest. Under the assumption that all interaction terms are null, the distribution of the p-values should be uniform and the points on the diagram should be on a line passing through the origin. For convenience, this line is represented (function `abline()`❺). It is obtained from the function `lm()`❸ and a specification that the line passes through the origin (`~ -1+`❹). It can be noted that in the present example, the empty dots are approximately on the line: the problem of interaction is globally negligible. The full dots correspond to an imaginary situation where there seems to be surprisingly small p-values for some interaction terms. These interaction terms should be investigated carefully.

What should be concluded from all these analyses? If the interaction between "direct" and "f.type.center" had actually been anticipated before the analysis, then the results are of interest: It appears that a low level of self-directedness globally increases the probability of being at high risk of suicide attempt (this is derived from the analysis without the interaction term); it also appears that this effect is small (and perhaps reversed) in high-security units, while it is more marked in the other types of prison. This result is the reverse of what was expected, and is therefore difficult to interpret. Perhaps life is particularly ritualized in high-security units so that the need for strong self-directedness could be less crucial in enduring imprisonment. In all events, this surprising result shows that it is difficult to predict an interaction between two variables, and that explanations can always be proposed a posteriori.

6.5 Assessing the Relative Importance of Predictors

In a few words: The ultimate aim of a model is often to determine which predictors are likely to have a stronger impact on the outcome. Because this determination is crucial, there is a large volume of literature on it. Unfortunately, this literature is particularly controversial (O'Grady 1982; Greenland, Schlesselman, and Criqui 1986; Greenland et al. 1991; Rothman and Greenland 1998; Menard 2004; Mood 2009).

A possible explanation of these controversies is that the notion of "strong impact" is polysemous and that its formal translation can vary from one discipline to another.

In epidemiology (for example), most models are logistic regressions that explain a binary variable (having a disease or not) by a series of binary risk factors. In this case, the exponential of the regression coefficients are adjusted odds-ratios (*aOR*), and they can be used to represent the impact of a given risk factor on the probability of having the disease. Indeed, when the prevalence of the disease is low, the probability of having the disease is approximately *aOR* times greater when the subjects have the risk factor than