Coding of contrasts and interactions in non-experimental research

Herbert Matschinger

University of Leipzig
Department of Psychiatry
e-mail: Herbert.Matschinger@medizin.uni-leipzig.de

1 Introduction

The following textbook is concerned with the estimation of effects from categorical and continuous variables and its interactions on **one** continuous variable in linear models. What will be said also holds, in principle, for non-linear probability models, although things become more difficult as far as predicted probabilities instead of logits are concerned. We will describe and explain several coding schemes with and without interactions and demonstrate how to interpret the model parameters. This guideline seems to be necessary, because information about how to construct design matrices adequately is not easy to be obtained. On the contrary, it is dotted about here and there in books and articles employing different notation, making it hard to understand for the non-technical reader and/or applied researcher. The main goal of this booklet therefore is to demonstrate the applicability of several of these schemes for non-experimental research. The focus is on the interpretation of the parameters resulting from a more or less complex design, employed as a matrix of predictors. Problems of statistical inference will only be allusively addressed since this is not the main topic of this small textbook. Nevertheless, the estimation of standard errors for combination of parameters will be addressed and outlined briefly.

1.1 A general overview

This guideline is an attempt to bridge the gap between what can be found in books and articles about Variance/Covariance analysis, and in texts on linear regression models. In most instances the focus is on testing, and the interpretation of more complex combination of parameters is badly neglected. We also want to show that the very common distinction between continuous and categorical predictors is entirely vacuous. For a linear model all the predictors are treated "as is", which means they are treated as vectors of real numbers, and it is up to the researcher to take care of what these figures actually portray. Or, the other way round, one must establish a set of vectors which decompose the dependent variable into a set of expected values such mapping the hypotheses under consideration.

Four different coding schemes will be presented and evaluated with suitable examples. Dummy coding, effect coding, forward adjacent coding and Helmert coding. In the final chapter a way to compute ones own contrasts is presented, first by means of how to define one of the already illustrated contrasts, and secondly by an arbitrary scheme not predefined by one of the contrasts available from the software used throughout this textbook. For sake of space we will not present more than one example for each of the contrast schemes, rather the reader is asked to repeat a particular example using the other dataset provided.

The reader should bear in mind that this text was not written to serve as a substitute for comprehensive and far more detailed books on linear models, like those cited in the text below. Important topics like orthogonality of a design and all the problems of multiple testing, for instance, will not be addressed at all. Additionally, problems arising from non-balanced designs are not treated. Several corrections outlined, for instance, by Pedhazur (Pedhazur, 1982) will not be described in the following pages. All this, and a lot of even more important topics can be found in the literature about ANOVA/ANCOVA (to mention only two: Rabe-Hesketh & Everitt, 2004; Rencher, 2000). However, it might serve as a compendium on how to construct and interpret more complex designs derived from theoretical assumptions.

1.2 Data

We will present the empirical examples using data from two different studies:

1. A study on attitudes towards the mentally ill was conducted in Germany in 2001 (N=5025). As a dependent variable we employ an index to measure the expected social distance towards a person described by means of a cases story. This index is the sum of 7 items regarding 7 situations, from neighbourhood to marriage. For each of them the respondent is asked to express his/her opinion concerning their willingness to accept the person described in the cases-story. The item-format is of the Likert type from 0 (definitely accept to 4 (definitely reject). The sum therefore covers a range from 0 to 28. This score will be called `dist1sum`. To identify the person, two different case stories are presented by means of a so called "vignette". Each respondent is presented only one of them, which results in a dichotomous variable called `q0rec` in the following. The first case-story gives a sketchy description of a schizophrenic episode (coded 0), the second one present the symptoms of a major depression (coded 1). Both stories can be found in the appendix. For a second categorical predictor we employ a variable, resulting from an open question, where respondent were asked to classify the "vignettes" just outlined above. Responses were classified into four categories: 1: The respondent classifies the story correctly either as schizophrenia or depression. 2: The respondent labelled the person described at least as mentally ill. 3: The person described is labelled aberrant or ill. 4: A residual

category comprising all other answers. We will employ different forms of collapsing to 3 categories or to a dichotomous variable. Variables will be called `skt` (4 cat.), `skt3` (3 cat.) or `sktdrec` (dichotomous).

Table 1: **`tab skt sktdrec`**

```
       |              skt3                   sktdrec
   skt |         1          2          3 |         0          1 |     Total
-------+---------------------------------+-   --------------------+----------
  1.00 |     1,511          0          0 |     1,511          0 |     1,511
  2.00 |     1,816          0          0 |     1,816          0 |     1,816
  3.00 |         0        995          0 |         0        995 |       995
  4.00 |         0          0        703 |         0        703 |       703
-------+---------------------------------+-   --------------------+----------
 Total |     3,327        995        703 |     3,327      1,698 |     5,025
```

2. The second study we adopt for presenting the examples is the ESEMED study in 6 countries (N=21425). We employ the mental health score (`mcs12`) of the SF36 as dependent variable. For predictors we use the 6 countries (coded by means of several different schemes) and the dichotomous partition of the sample into non-depressive and depressive respondents respectively by means of `dsm_mde`.

We will not go into further details regarding the goals of these studies, nor will we describe or critically evaluate the characteristics of samples and variables. All interpretations of results should be considered cautiously and do not claim exclusivity. The reader is encouraged to apply the following examples to his/her own problems with his/her own data sets. Data sets of the two studies will be provided.

1.3 Software tools

Although each contrast for virtually all the models may be estimated by any statistical package, we will focus on STATA (StataCorp, 2005)and one of the pre-commands for STATA named `xi3`. Furthermore we will make use of the command `estimate`s and its options in order to store several models in memory and compare them by means of likelihood ratio tests (`lrtest`). In order to test sets of parameters or linear combination of parameters, we will make use of the command `test` or `lincom` respectively. We also will make use of the matrix facilities of STATA, particularly to estimate the standard errors of parameters not immediately provided by a particular regression model. To compute and draw conditional and partial effects (slopes) we also employ the program `postgr3`, a post estimation command which only works in conjunction with the `spostado` suite provided by Jeremy Freese and Scott Long (Long & Freese, 2003). STATA user can implement these tools for free by typing:

```
net from http://www.indiana.edu/~jslsoc/stata
net install spostado
```

Most of the figure and table captions will provide the necessary STATA commands by which the graph models were generated.

2 Dummy coding

By "dummy coding" we mean that each level of a categorical variable will be compared to a reference category level (also called "omitted category"). Variables or indicator variables should be coded zero and one, since then the effect parameter displays the exact mean difference between the omitted category and the category coded 1. Employing a predictor coded zero and one, and thereby distinguishing between two groups of observations is equivalent to a t-test. If the variables are coded, for instance, 0 and 2, the actual mean difference is twice the regression parameter, and the t-value remains the same, because the standard error is now also half of the standard error of the (0,1) coded predictor. If the variable is coded 1 and 2 things will become more complicated and will not be addressed further more. To cut the long story short: To estimate mean differences by dummy coding, always adopt the (0,1) coding scheme. Nevertheless we do present the results of both codings below. In the first example we employ `dist1sum` as dependent variable and `sktdrec` as predictor. The goal of this analysis is to find out whether the two subgroups differ with respect to their means on the social distance scale.

2.1 One dichotomous predictor

Table 2: **`regress dist1sum sktdrec`**

```
      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  1,  4974) =   48.49
       Model |  2049.24049     1  2049.24049           Prob > F      =  0.0000
    Residual |  210214.383  4974  42.2626424           R-squared     =  0.0097
-------------+------------------------------           Adj R-squared =  0.0095
       Total |  212263.624  4975   42.666055           Root MSE      =   6.501

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     sktdrec |  -1.355258   .1946273    -6.96   0.000    -1.736813   -.9737026
       _cons |   17.25342    .113391   152.16   0.000     17.03113    17.47572
```

The tabulation of means for both categories of `sktdrec` shows, that we obtain a test for the differences of means. A t-test will, of course, provide exactly the same results. Since we have only one predictor the square root of the F(1,4974)- value is exactly the t-value.

Table 3: **`tab sktdrec,sum(dist1sum)`**

```
            |          Summary of egen
            |     dist1sum=rsum(q15a-q15g) if
            |             distmiss==0
    sktdrec |        Mean   Std. Dev.       Freq.
------------+------------------------------------
          0 |   17.253423   6.4038822        3287
          1 |   15.898165   6.6859336        1689
------------+------------------------------------
      Total |   16.793408   6.5319258        4976
```

Coding the predictor 0 and 2 instead of 0 and 1, we obtain the same value for the constant (the omitted category) but the effect (now called `sktdrec02`) is only half of the actual mean difference we already know from Table 2 (compare the figures in bold face). This is necessarily so, since the regression equation reads:

ŷ= `_cons + sktdrec02*0` for the omitted category, and

ŷ= `_cons + sktdrec02*`***2*** for the second category.

Table 4: **`regress dist1sum  sktdrec02`**

```
      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  1,  4974) =   48.49
       Model |  2049.24049     1  2049.24049           Prob > F      =  0.0000
    Residual |  210214.383  4974  42.2626424           R-squared     =  0.0097
-------------+------------------------------           Adj R-squared =  0.0095
       Total |  212263.624  4975   42.666055           Root MSE      =   6.501

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   sktdrec02 |   -.677629   .0973137    -6.96   0.000    -.8684067   -.4868513
       _cons |   17.25342    .113391   152.16   0.000     17.03113    17.47572
------------------------------------------------------------------------------
```

The real mean difference is now two times the parameter, the standard error of the regression parameter is exactly half of what is observed if 0/1 coding is adopted. Therefore inference remains invariant; the constant also still represents the mean with respect to the omitted category (definition of to be mentally ill). Further on we will never use such a coding system; it simply serves to justify the usual 0/1 coding to estimate simple mean differences. Of course, this scheme holds even for predictors with more categories, since the multicategory variable is then decomposed into several dichotomous variables, as will be shown in section 2.3.

Study suggestions 1: Try to estimate the regression coding the predictor 2 & 4. Interpret both the effect parameter and the constant.

2.2 Two dichotomous predictors

For the next example we again use the social distance score (`dist1sum`) as the dependent continuous variable. The type of the vignette questionnaire `q0` and `skt` (4 categories of labelling) is used as independent variables. This variable discriminates between those respondents who label the person described in the vignette at least as mentally ill, and those who did not (comp. Section 1.2.). With the STATA command: `tab sktdrec q0rec, sum(dist1sum)` you will obtain a 2x2 table displaying the means of the dependent variable (`dist1sum`) within the different levels of the independent variables.

Table 5: Distance to people with mental disorders (`dist1sum`, mean scores) by `q0rec` and `sktdrec`

	q0rec		Total
Sktdrec	0	1	
0	y_{11} 19.200808	y_{12} 15.081725	17.253423
1	y_{21} 17.116022	y_{22} 14.984456	15.898165
Total	18.586488	15.044462	16.793408

By means of the command: `regress dist1sum q0rec sktdrec` we obtain the following regression coefficients (Table 6) and the estimated means for each of the four cells (Table 7). Results are stored under the name `mo1`.

Table 6: `regress dist1sum q0rec sktdrec`

```
      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  2,  4973) =  212.94
       Model |  16744.0457     2  8372.02287           Prob > F      =  0.0000
    Residual |  195519.578  4973  39.3162232           R-squared     =  0.0789
-------------+------------------------------           Adj R-squared =  0.0785
       Total |  212263.624  4975   42.666055           Root MSE      =  6.2703

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       q0rec |  -3.452283   .1785708   -19.33   0.000     -3.80236   -3.102205
     sktdrec |  -1.014958   .1885438    -5.38   0.000    -1.384587   -.6453289
       _cons |   18.88556   .1381608   136.69   0.000     18.61471    19.15642
------------------------------------------------------------------------------
```

```
predict dist1sump if e(sample)
(option xb assumed; fitted values)
(49 missing values generated)
```

```
estimates store mo1
```

Table 7: `tab sktdrec q0rec,sum(dist1sump)`

```
           |        q0rec
   sktdrec |         0          1 |     Total
-----------+----------------------+----------
         0 | 18.885563  15.433281 | 17.253422
           |         .          0 | 1.7238418
           |      1733       1554 |      3287
-----------+----------------------+----------
         1 | 17.870605  14.418323 | 15.898164
           |         0          . | 1.7089851
           |       724        965 |      1689
-----------+----------------------+----------
     Total | 18.586487  15.044462 | 16.793408
           | .46280669  .49350943 | 1.8345671
           |      2457       2519 |      4976
```

If this model would suffice, the mean differences between the categories of one predictor are equal for both categories of the other:

1. y_{11} - y_{12} | X2 == 0
2. y_{21} - y_{22} | X2 == 1 = $\mathbf{b_1}$

in terms of the additive linear combination of parameters :

$$(c + b_1*0 + b_2*0) - (c + b_1*1 + b_2*0) = \mathbf{-b_1}$$
$$(c + b_1*0 + b_2*1) - (c + b_1*1 + b_2*1) = \mathbf{-b_1}$$

Of course, this also holds for X1.

The model above is not saturated – the means for the 4 cells are estimated by only 3 parameters – and the estimated means do not equal the real means (Table 5 and Table 7). Therefore we will immediately turn to the fully saturated model, which model not only the differences for each predictor, but also the dependence of each effect on the different values of the other predictor. Here we use a pre-command feature of STATA: `xi3` for the first time, which helps to built special contrasts and interactions. We did not employ this feature up to now, since both predictors are already coded 0/1. To get an interaction term, a simple star is inserted. If this command (`xi3`) is not available the interaction variable has to be computed beforehand by simply setting up the product of `q0rec` and `sktdrec`, generating a new variable. This model is stored under the name `mo2`.

Table 8: **`xi3:regress dist1sum q0rec*sktdrec`**

```
      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  3,  4972) =  151.98
       Model |   17829.863     3  5943.28765           Prob > F      =  0.0000
    Residual |  194433.761  4972  39.1057443           R-squared     =  0.0840
-------------+------------------------------           Adj R-squared =  0.0834
       Total |  212263.624  4975   42.666055           Root MSE      =  6.2535

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       q0rec |  -4.119083   .2184718   -18.85   0.000    -4.547384   -3.690782
     sktdrec |  -2.084786   .2767288    -7.53   0.000    -2.627296   -1.542275
     _Iq0Xsk |   1.987517   .3771836     5.27   0.000     1.248071    2.726963
       _cons |   19.20081   .1502177   127.82   0.000     18.90631     19.4953
------------------------------------------------------------------------------
```

`estimates store mo2`

To calculate the means of the dependent variable (Table 5), one has to use the equation for the linear regression model and the regression coefficients from the STATA output of the

`ŷ = cons + b1*q0rec + b2*sktdrec + b3*(q0rec*sktdrec)`

The precommand `xi3` will produce a variable called `_Iq0Xsk` which represents the product of `q0rec` and `sktdrec`. We should have a closer look at how the cell means are generated by linear combinations of the model parameters. If, for instance, you insert into the equation the following values you will get:

y_{11} = cons + b1 (q0rec=0) + b2 (sktdrec=0) + b3 (q0rec=0*sktdrec=0) =

cons + b1*0 + b2*0 + b3*0= **cons** = **19.200808**

This cell represents the constant in the regression model, since both independent variables are coded zero (`sktdrec=0` and `q0rec=0`). Therefore, this cell mean represents the mean for **both** reference categories.

y_{21} = cons + b1 (q0rec=0) + b2 (sktdrec=1) + b3 (q0rec=0*sktdrec=1) =
cons + b1*0 + b2*1 + b3*0 = **cons + b2**

y_{21} = 19.200808 + (-2.084786) = **17.116022**

The mean in this cell represents the mean for people with q0rec = 0 and sktdrec =1. To obtain this cell mean, you have to add the difference between both levels of `sktdrec` (represented by the coefficient of `sktdrec` = -2.084786) to the constant.

y_{12} = cons + b1 (q0rec=1) + b2 (sktdrec=0) + b3 (q0rec=1*sktdrec=0) =
cons + b1*1 + b2* 0+ b3*0= **cons + b1**

y_{12} = 19.200808 + (-4.119083) = **15.081725**

The mean in this cell represents the mean for people with `q0rec = 1` and `sktdrec =0`. To obtain this cell mean you have to add the difference between both levels of q0rec (represented by the coefficient of `q0rec` = -4.119083) to the constant.

y_{22} = cons + b1 (q0rec=1) + b2 (sktdrec=1) + b3 (q0rec=1*sktdrec=1) =
cons + b1*1 + b2* 1+ b3*1= **cons + b1 + b2 + b3**

y_{22}= 19.200808 +(-4.119083) + (-2.084786) + 1.987517 = **14.984456**

The mean in this cell represents the mean for people with `q0rec` = 1 and `sktdrec` =1. To obtain this cell mean you have to add the difference between both levels of q0rec (represented by the coefficient of `q0rec` = -4.119083) and the difference of both levels of `sktdrec` (represented by the coefficient of `sktdrec` = -2.084786) as well as the coefficient of the interaction term (`_Iq0Xsk` = 1.987517) to the constant. In Table 9 the more general way of reconstructing the cell means from a combination of the 4 parameters is demonstrated.

Table 9: Decomposition table

	q0rec	
sktdrec	**0**	**1**
0	**cons** y_{11}	**cons + b1** y_{12}
1	**cons + b2** y_{21}	**cons + b1 + b2 + b3** y_{22}

In the case of a saturated model with interaction (Table 8), the so called “simple slopes” (printed in bold face) are always additive linear combinations of the following form:

ŷ = cons + (**b1 + b3 sktdrec**) q0rec +b2sktdrec or:

ŷ = cons + (**b2 + b3 q0rec**) sktdrec +b1sktdrec

We see that the simple effects depend on both the value of the interaction term and the value of the predictor which is weighted by the interaction term (here b3). The coefficients b1 or b2 portray the effect of q0rec or sktdrec only if the other variable **or the coefficient b3** is zero. This simple relation always applies, whether the predictors are categorical (dichotomous or multicategorical) and therefore coded in one or the other way, or continuous, where we have to take care about the **meaning** of zero. Of course, we could also carry out a likelihood ratio test (lrtest) to compare the restricted model with the fully saturated one. Taking the square root of the chi^2 value 27.71, we get 5.264, which is, apart from rounding errors, the t-value from the regression analysis, found in Table 8. Strictly speaking, this lrtest is not necessary if only one parameter is tested. To administer this test, the two models are stored in memory immediately after the estimation by the command estimates store [modelname]. For [modelname] we adopted, as said before, mo1 and mo2.

Table 10: **lrtest mo1 mo2**

```
Likelihood-ratio test                                 LR chi2(1)  =     27.71
(Assumption: mo1 nested in mo2)                       Prob > chi2 =    0.0000
```

2.3 Two predictors (one dichotomous, one three categorical variable)

Now we employ the dichotomous predictor (q0rec) and skt3 as independent variables to estimate the dist1sum score. We collapsed the first two categories of skt (correct labelling and mentally ill) in order to get a 3-categorical variable. Let's first have a look at the 3x2 table, displaying the means of the dependent variable (dist1sum) for the different levels of the independent variables.

Table 11: **tab skt3 q0rec,sum(dist1sum)**

	q0rec		Total
skt3	0	1	
1	19.200808 y_{11}	15.081724 y_{12}	17.253423
2	16.798761 y_{21}	14.694611 y_{22}	15.380424
3	17.371571 y_{31}	15.636364 y_{32}	16.633237
Total	18.586488	15.044462	16.793408

Since the categorical variable `skt3` has three levels, we will have to use two indicator variables (`I1` and `I2`) to represent the categories of `skt3`. In all of the coding systems presented you need to create k-1 variables (where k represents the number of levels of the categorical variable). The remaining level of the categorical variable is redundant. Again, we do not need a new variable for representing the case-story, since it is already coded 0 and 1.

Table 12: Indicator variables for `skt3` if category 3 serves as reference category:

skt3	I1	I2
1	1	0
2	0	1
3	0	0

Using the last category as reference, of course, is an arbitrary decision and the reader is invited to estimate and interpret models with other "omitted" categories. The following regression analysis (Table 13) shows the parameter estimates of a model without interaction.

Table 13: `xi3:regress dist1sum q0rec i.skt3`

```
i.skt3            _Iskt3_1-3          (naturally coded; _Iskt3_3 omitted)

      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  3,  4972) =  142.53
       Model |  16809.1834     3  5603.06113           Prob > F      =  0.0000
    Residual |   195454.44  4972  39.3110298           R-squared     =  0.0792
-------------+------------------------------           Adj R-squared =  0.0786
       Total |  212263.624  4975   42.666055           Root MSE      =  6.2699

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       q0rec |  -3.419012   .1804199   -18.95   0.000    -3.772715    -3.06531
    _Iskt3_1 |   .7818018   .2614418     2.99   0.003     .2692605    1.294343
    _Iskt3_2 |  -.4029666   .3130475    -1.29   0.198    -1.016678    .2107446
       _cons |   18.08803   .2494253    72.52   0.000     17.59905    18.57702
------------------------------------------------------------------------------
```

```
estimates store mo1

predict dist1sump3 if e(sample)
(option xb assumed; fitted values)
(49 missing values generated)
```

The estimated means of the criterion, assuming that all the mean differences with respect to the omitted category are equal for both values of `q0rec` are provided in Table 14.

Table 14: **`tab skt3 q0,sum(dist1sump3)`**

```
        Means, Standard Deviations and Frequencies of Fitted values

           |
      skt3 |  Schizo      Depress |     Total
-----------+----------------------+----------
         1 | 18.869835  15.450822 | 17.253423
           |         .  2.771e-06 | 1.7072295
           |      1733       1554 |      3287
-----------+----------------------+----------
         2 | 17.685066  14.266053 | 15.380424
           |         0          . | 1.6033776
           |       323        668 |       991
-----------+----------------------+----------
         3 | 18.088032   14.66902 | 16.633237
           |         .          . | 1.6916361
           |       401        297 |       698
-----------+----------------------+----------
     Total | 18.586488  15.044462 | 16.793408
           | .45175368     .52847 | 1.8381326
           |      2457       2519 |      4976
```

If we compare the estimated means of this non-saturated model with the actual means, we see immediately that these figures differ considerably. However, we should interpret the regression parameter `q0rec` as the difference for social distance between the two case-stories, and this difference must be the same for all categories of `skt3`, since we have no interaction. This relation must also hold for the two dummy variables `_Iskt3_1` and `_Iskt3_2`: They represent the difference between the omitted category (3) and one of the other cells. These differences are the same for both columns. But, as said before, the predicted means are quite different from the actual means, and it becomes necessary to estimate the fully saturated model, including two additional (interaction-) effects in order to parameterise the conditional differences for category 1 & 2 of `skt3` with regard to `q0rec`. Of course, the estimated means will then be equal to the real means of each cell, just as in section 2.2. We obtain the following results using the pre-command `xi3` (Table 15).

Both the model without and with interaction effects have been stored with the names `mo1` and `mo2` respectively. The likelihood ratio test clearly shows that the interaction effects are statistically indispensable and we should only accept the saturated model. To calculate the means of the dependent variable shown in Table 11, you have to use the equation (including the indicator variables) for the linear regression model and the obtained coefficient from the STATA output of the regress command.

Table 15: **`xi3:regress dist1sum q0rec*i.skt3`**

```
i.skt3            _Iskt3_1-3         (naturally coded; _Iskt3_3 omitted)

      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  5,  4970) =   92.50
       Model |  18070.9007     5  3614.18013           Prob > F      =  0.0000
    Residual |  194192.723  4970  39.0729825           R-squared     =  0.0851
-------------+------------------------------           Adj R-squared =  0.0842
       Total |  212263.624  4975   42.666055           Root MSE      =  6.2508

------------------------------------------------------------------------------
      dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
b1       q0rec |  -1.735207   .4785371    -3.63   0.000    -2.673351   -.7970636
b2(I1)_Iskt3_1 |   1.829237   .3463889     5.28   0.000     1.150162    2.508312
b3(I2)_Iskt3_2 |  -.5728095   .4673413    -1.23   0.220    -1.489005    .3433857
b4    _Iq0Xsk1 |  -2.383876   .5260111    -4.53   0.000     -3.41509   -1.352662
b5    _Iq0Xsk2 |  -.3689434   .6391081    -0.58   0.564    -1.621877    .8839906
         _cons |   17.37157    .312152    55.65   0.000     16.75962    17.98353
------------------------------------------------------------------------------
```

`estimates store m2`

`lrtest mo1 mo2`

```
Likelihood-ratio test                         LR chi2(2)  =      32.23
(Assumption: mo1 nested in mo2)               Prob > chi2 =     0.0000
```

In this very case we do not test one single parameter, but rather compare two models where the saturated model has two more parameters (`b4` and `b5`). Therefore the likelihood ratio test is the adequate way to decide between the restricted and the saturate model. Again, we should have a closer look at how each cell can be represented by an additive linear combination of the parameters. This will help decide which parameter (or combination of parameters) should be tested for which hypotheses.

```
ŷ= cons + b1*q0rec + b2*I1 + b3*I2 + b4*(q0rec*I1) + b5*(q0rec*I2)
                                         _Iq0Xsk1        _Iq0Xsk2
```

If you insert the corresponding values, you will get the following results:

y_{11} = cons + b1 (q0rec=0) + b2 I1 (I1=1) + b3 I2 (I2=0) + b4 (q0rec=0*I1=1) + b5 (q0rec = 0 * I2 = 0)= cons + b1*0 + b2*1 + b3*0 + b4*0 + b5*0 =
cons + b2
y_{11} = 17.371571 + 1.829237 = **19.200808**

y_{21} = cons + b1 (q0rec=0) + b2 I1 (I1=0) + b3 I2 (I2=1) + b4 (q0rec=0*I1=0) + b5 (q0rec = 0 * I2 = 1) = cons + b1*0 + b2*0 + b3*1 + b4*0 + b5*0 =
cons + b3
y_{21} = 17.371571 + (-0.5728095) = **16.798761**

y_{31} = cons + b1 (q0rec=0) + b2 I1 (I1=0) + b3 I2 (I2=0) + b4 (q0rec=0*I1=0) + b5 (q0rec = 0 * I2 = 0) = cons + b1*0 + b2*0 + b3* 0 + b4*0 + b5*0 =
cons
y_{31} = **17.371571**

y_{12} = cons + b1 (q0rec=1) + b2 I1 (I1=1) + b3 I2 (I2=0) + b4 (q0rec=1*I1=1) + b5 (q0rec = 1 * I2 = 0) = cons + b1*1 + b2*1 + b3* 0 + b4*1 + b5*0 =
cons + b1 + b2 + b4
y_{12} = 17.371571 + (-1.735207) + 1.829237 + (-2.383876) = **15.081725**

y_{22} = cons + b1 (q0rec=1) + b2 I1 (I1=0) + b3 I2 (I2=1) + b4 (q0rec=1*I1=0) + b5 (q0rec = 1 * I2 = 1) = cons + b1*1 + b2*0 + b3* 1 + b4*0 + b5*1 =
cons + b1 + b3 + b5
y_{22}= 17.371571 + (-1.735207) + (-0.5728095) + (-0.3689434) = **14.694611**

y_{32} = cons + b1 (q0rec=1) + b2 I1 (I1=0) + b3 I2 (I2=0) + b4 (q0rec=1*I1=0) + b5 (q0rec = 1 * I2 = 0) = cons + b1*1 + b2*0 + b3* 0 + b4*0 + b5*0 =
cons + b1
y_{32}= 17.371571 + (-1.735207) = **15.636364**

Table 16 shows in a more comprehensive way, how each of the 6 cells means (comp. Table 11) can be represented by an additive linear combination of the model parameters.

Table 16: Decomposition table

	q0rec	
skt3	**0**	**1**
1	y_{11} **cons + b2**	y_{12} **cons + b1 + b2 + b4**
2	y_{21} **cons + b3**	y_{22} **cons + b1 + b3 + b5**
3	y_{31} **cons**	y_{32} **cons + b1**

The additive linear combinations (in bold face) provide the information on how to interpret the regression parameters. The non-saturated model simply assumes that both b4 and b5 are zero, which does not hold, as the lrtest clearly has shown. Now, what is called the simple slopes or conditional effect of the variable q0rec and both dummy variables _Iskt3_1 and _Iskt3_2 are not "main effects" any more. q0rec is the difference between the two groups indicated by this variable, but only for the respondents in category 3 (omitted category) of skt3. To obtain the difference between the 2 case stories conditioned on category 2 of skt3 one has to add b5, because:

$\Delta(y_{21}, y_{22})$ = (cons + b3) − (cons + b1 + b3 + b5) = **b1+b5**

Since b5 (-.3689434) is small and not significant we can assume the effect of q0rec does not differ between category 2 and 3 of skt3; it remains to be b1.

$\Delta(y_{11}, y_{12})$ = (cons + b2) − (cons + b1 + b2 + b4) = **b1+b4**

For the first category of skt3 this difference is b4 (-2.383876) and highly significant. We should conclude that the difference between q0rec categories (vignettes) conditioned on category 1 of skt3 (labelled as mentally ill) is considerably larger than conditioned on category 3. Results are not surprising: Category 2 and 3 of skt3 are rather similar. There is not much impact on the effect of the case story. But for those who labelled the person described in the vignette as *mentally ill*, being presented with a schizophrenic episode or a major depression, is of great importance for their scoring on the social distance scale. The difference is more than 4 points, on a scale from 0 to 28 (b1 + b4). Compare the first row of Table 16.

2.4 Two predictors (one dichotomous and one continuous)

In this chapter we will demonstrate how a categorical variable, one continuous variable and their interaction may be implemented as predictors and how the parameters have to be interpreted. This chapter should be considered only as a short outline, further reading is necessary to become genuinely familiar with all the problems (Aiken & West, 1991). For this example we additionally plug in the age of the respondents as a predictor of social distance. Everything said about "simple slopes" in section 2.2 still holds. If an interaction is adopted as a predictor, which is simply the product of at least two predictors, the conditional effects are always the effects of one of the predictors, given the other predictor is zero. Therefore, the researcher has to know what "zero" actually means. In the case of a dummy coded categorical variable, this is quite simple and straightforward. For a continuous variable we have to take care that a value of zero is both in the range of the observable values and represents a value where the conditional effect for the categorical variable is really of interest. If we employ the age "as is", both criteria are obviously violated. Observations with an age of zero definitely are not a part of the sample **and** the population, and, of course, any test regarding this "group" of observations makes no sense at all. Usually, to solve this problem, continuous variables are centred at their (sample) mean, so the conditional effect is the effect at the mean of the continuous variable. We first estimate a model without and then with an interaction between `agec` (age centred at its sample mean) and `q0rec`, which can be found in Table 17 and Table 18 respectively.

Table 17: **`xi3:regress dist1sum q0rec agec`** **(age "mean centred")**

```
      Source |       SS       df       MS              Number of obs =    4938
-------------+------------------------------           F(  2,  4935) =  224.78
       Model |  17572.1465     2  8786.07327           Prob > F      =  0.0000
    Residual |  192896.307  4935  39.0873976           R-squared     =  0.0835
-------------+------------------------------           Adj R-squared =  0.0831
       Total |  210468.454  4937  42.6308393           Root MSE      =   6.252

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       q0rec |  -3.551218   .1779552   -19.96   0.000     -3.90009   -3.202347
        agec |   .0376527   .0051973     7.24   0.000     .0274638    .0478416
       _cons |    18.5861    .126621   146.79   0.000     18.33787    18.83434
------------------------------------------------------------------------------
```

`Estimates store q0recage1`

Table 18: **xi3:regress dist1sum q0rec*agec (with interaction)**

```
      Source |       SS       df       MS              Number of obs =    4938
-------------+------------------------------           F(  3,  4934) =  152.54
       Model |  17863.9569     3  5954.65228           Prob > F      =  0.0000
    Residual |  192604.497  4934  39.0361769           R-squared     =  0.0849
-------------+------------------------------           Adj R-squared =  0.0843
       Total |  210468.454  4937  42.6308393           Root MSE      =  6.2479

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       q0rec |  -3.550065    .177839   -19.96   0.000    -3.898708   -3.201421
        agec |   .0231622   .0074206     3.12   0.002     .0086146    .0377098
    _Iq0Xag |   .0284069   .0103898     2.73   0.006     .0080383    .0487756
       _cons |   18.58454   .1265393   146.87   0.000     18.33646    18.83261
------------------------------------------------------------------------------
```

`Estimates store q0recage2`

To facilitate the comparison between the two hierarchically nested models, we employ another feature of the `estimates` command. Since the names of the models only differ with respect to the last character (1 or 2), we can shorten the command by using the wildcard “*”:

Table 19: estimates table q0recage*,star stats(N r2 ll) b(%9.2f)

```
--------------------------------------------
    Variable |  q0recage1      q0recage2
-------------+------------------------------
       q0rec |     -3.55***       -3.55***
        agec |      0.04***        0.02**
     _Iq0Xag |                     0.03**
       _cons |     18.59***       18.58***
-------------+------------------------------
           N |   4938.00        4938.00
          r2 |      0.08           0.08
          ll | -16056.08      -16052.34
--------------------------------------------
    legend: * p<0.05; ** p<0.01; *** p<0.001
```

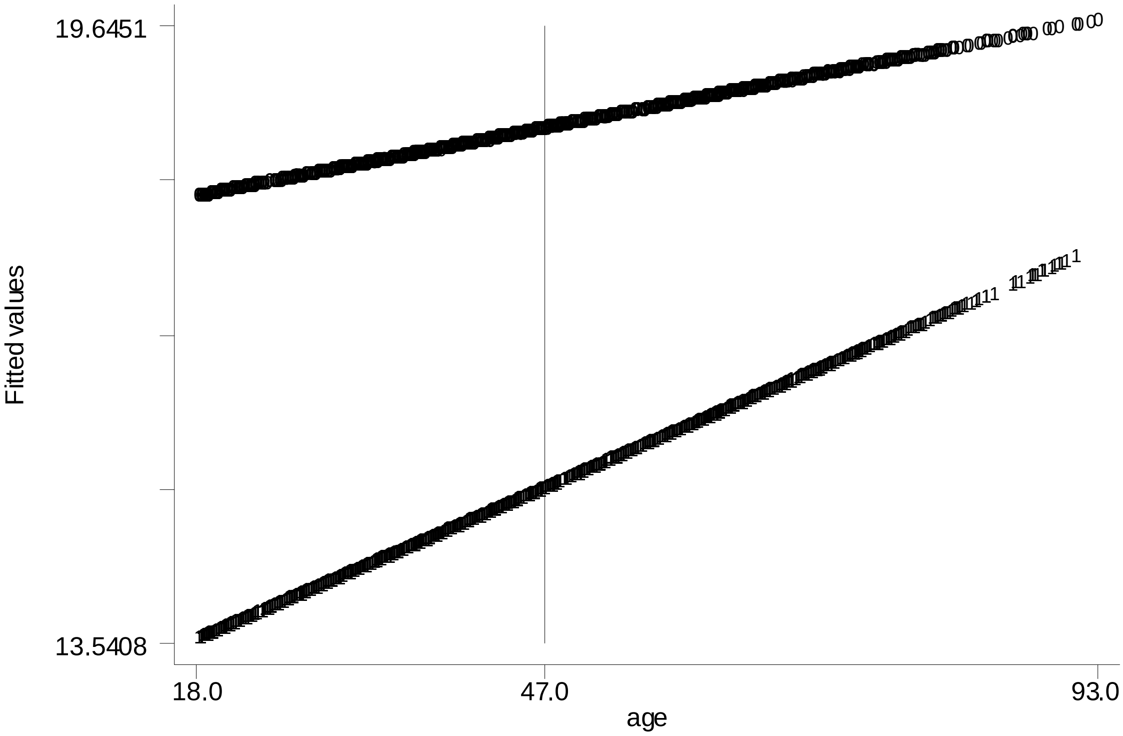
We plot the predicted values of the second model against age and adopt the codes of `q0rec` for a plotting symbol. As said before, we employ the mean centred age (`agez`) as predictor for reasons presented above. The conditional effect of `q0rec` exactly represents and tests the difference in social distance for this value of age. Two important facts should not be overlooked:

1. The mean of age should be the mean age for only those observations which are used to estimate the regression.
2. The test of the conditional effect (here `q0rec`) is a test at the mean age of the ***sample.*** Of course, this value may differ from the mean of the population, which is not taken into account. However, for the purpose of this memo this might be of minor importance.

To get the predicted values for the saturated model, we type `predict distage2 if e(sample)` and plot the age against these variable.

Figure 1: **Scatterplot of predicted values against the non- centred age (mean age 47)**

```
gr7 distage2 age,s([q0rec]) xlab(18,47,93) xline(47)
```



We see that although the R^2 does not change at all, the likelihood test (and the interaction term) turned out to be significant. Here we are confronted with a very important problem of how to interpret model results: The difference between statistical and practical significance. Since the two lines presenting the linear fit for each of the two groups (`q0rec(0,1)`), the conditional "slope" changes, if the value zero for the continuous variable changes its meaning. Of course, we could have centred the age variable for other values: for instance at 18 (youngest) and 90 (oldest), and it would be possible to centre `age` at 200, which will yield an out of range estimation for the conditional effect `q0rec`: 0.79, but this is not significant. If only our respondents would have reached an age of 200 years, the two case stories would produce no more difference at least with respect to social distance.

Of course one can get both the conditional slopes of `q0rec` and their standard errors from the regression with a non-centred age. First we have a look at Table 20 below:

Table 20: Model age (without any centring) xi3:regress dist1sum q0rec*age

```
      Source |       SS       df       MS              Number of obs =    4938
-------------+------------------------------           F(  3,  4934) =  152.54
       Model |  17863.9569     3  5954.65229           Prob > F      =  0.0000
    Residual |  192604.497  4934  39.0361769           R-squared     =  0.0849
-------------+------------------------------           Adj R-squared =  0.0843
       Total |  210468.454  4937  42.6308393           Root MSE      =  6.2479

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       q0rec |   -4.89123   .5213757    -9.38   0.000    -5.913359   -3.869102
         age |   .0231622   .0074206     3.12   0.002     .0086146    .0377098
     _Iq0Xag |   .0284069   .0103898     2.73   0.006     .0080383    .0487756
       _cons |   17.49099   .3717408    47.05   0.000     16.76221    18.21977
------------------------------------------------------------------------------
```

We see that only the conditional slope for `q0rec` changes: as expected, it is then even bigger in absolute value for the 18-centred age, but, as said before, this value is formally correct but meaningless. To get the "simple slope" for the 18 year old respondents, we briefly reconsider the linear equation already presented after Table 9, and rewrite them for the very problem outlined in this section.

ŷ = cons + (**b1 + b3*age**)*q0rec +b2*age or:

ŷ = cons + (**b2 + b3*q0rec**)*age +b1*q0rec

The "simple slopes" are again printed in bold face, and their values depend on the value of `age` or `q0rec`. Here we we will only evaluate the first equation (`b1 + b3*age`) because the second slope results in only two values : it is `b2` if `q0rec` is zero or `b2+b3` if `q0rec` is one. To estimate the coefficient for `q0rec` if the age is 18, the mean age, or 90, we compute using the `display` command:

```
display (-4.89123 + (.0284069*18)) =                 -4.379906
display (-4.89123 + (.0284069*47.17427))  =          -3.551155
display (-4.89123 + (.0284069*90)) =                 -2.334609
```

Unfortunately, we do not currently know the standard errors of the simple slopes, but the equations above tell us that each of the coefficients is a weighted linear combination of coefficients for which the Variance/Covariance matrix is available from the analysis presented earlier (Table 20). Therefore the standard error of the simple slope is the square root of a weighted linear combination. We first store the variance/covariance matrix immediately after the regression analysis in a matrix arbitrarily called `ageV` by means of the command `matrix ageV = e(V).`

Table 21: `matrix list ageV` **(only subdiagonal displayed)**

```
symmetric ageV[4,4]
              q0rec          age      _Iq0Xag        _cons
  q0rec   .27183263
    age    .0025938   .00005506
_Iq0Xag  -.00509213  -.00005506    .00010795
  _cons  -.13819122   -.0025938     .0025938    .13819122
```

If we want to estimate the variance of the linear combination for age 18, we first have to estimate the following combination of variances and covariances, which can be defined by looking at the square of the weighted linear combination of parameters:

```
(b1   +   b3*18)* (b1   +   b3*18)
```

To obtain the combination of parameters necessary to estimate the variance of the simple slope we compute:

Equation 1:

$$s_{11} + s_{31}*18 + s_{31}*18 + s_{33}*18*18 =$$

$$s_{11} + 2*s_{31}*18 + s_{33}*18^2$$

The square root of this expression then is the standard error of the respective coefficient. If we employ the display command and the row and column indices of the matrix `age`, we simply have to type:

```
display sqrt(ageV[1,1]+(2*ageV[3,1]*18) +(ageV[3,3]*(18*18)))
```

to obtain the expected standard error **.35141301.** If we would have estimated the model with age centred at 18, we would have obtained:

Table 22: **Model age centred at 18**

```
------------------------------------------------------------------------------------

      Source |       SS       df       MS              Number of obs =    4938
-------------+------------------------------           F(  3,  4934) =  152.54
       Model |  17863.9569     3  5954.65229           Prob > F      =  0.0000
    Residual |  192604.497  4934  39.0361769           R-squared     =  0.0849
-------------+------------------------------           Adj R-squared =  0.0843
       Total |  210468.454  4937  42.6308393           Root MSE      =  6.2479

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       q0rec |  -4.379906    .351413   -12.46   0.000    -5.068831    -3.69098
       age18 |   .0231622   .0074206     3.12   0.002     .0086146    .0377098
     _Iq0Xag |   .0284069   .0103898     2.73   0.006     .0080383    .0487756
       _cons |   17.90791   .2503108    71.54   0.000     17.41719    18.39863
------------------------------------------------------------------------------
```

Estimating the model several times, with differently centred continuous predictors is certainly the simpler way to the desired results. However it is worth knowing, how these coefficients and standard errors are related. This holds for each possible (and impossible) value of the continuous predictor. If instead of 18, 90 is plugged into the equation, we will obtain the standard error of `q0rec` for the 90 year old respondents. As expected the effects for the differently centred age variables and the respective interaction term remains invariant, but the conditional effect of `q0rec` changes from –4.3 to –2.3. The difference in social distance attributed to the case story becomes smaller if only the age of the respondents increases (Figure 1). We also see, that this difference remains significant for all observed values of age.

Table 23: **`estimates table age18 agec age90,star stats(N r2) b(%9.2f)`**

```
-----------------------------------------------------------
    Variable |     age18           agec          age90
-------------+---------------------------------------------
       q0rec |      -4.38***       -3.55***       -2.33***
       age18 |       0.02**
     _Iq0Xag |       0.03**         0.03**         0.03**
        agec |                      0.02**
       age90 |                                     0.02**
       _cons |      17.91***       18.58***       19.58***
-------------+---------------------------------------------
           N |    4938.00        4938.00        4938.00
          r2 |       0.08           0.08           0.08
-----------------------------------------------------------
legend: * p<0.05; ** p<0.01; *** p<0.001
```

The more general algorithm to obtain the variances of linear combination of parameters needs a selection (row) vector `w` and the var/cov matrix already introduced above, which is called `ageV` in the following. The vector `w` selects and weights the coefficients from the matrix `ageV`. The selection vector has as many as elements as the `ageV` has rows (or columns). Since we need the square of combination we need to premultiply the matrix `ageV` with `w`, and to postmultiply with `w'`, and we get:

Equation 2: $s_{comb} = w*ageV*w'$ (compare, for instance, Aiken & West 1991:24pp)

The square root of s_{comb} is the standard error of the coefficient. How must the weight vector `w` be defined ? For the example above we need two parameters: `b1` represented by `q0rec`,and `b3` represented by `_Iq0Xag` in Table 21. Since we want to estimate the variance for an age of 18, the 2nd parameter need to be weighted by 18 (comp. Equation 1). Therefore, the row vector `w` has to be defined `w = [1 0 18 0]`. The 2nd scalar is zero, since we do not need the variance or covariance of any term with `age`: s_{22}. Since the constant also is not of interest, it must be weighted by zero either.

Equation 3:

$$[1\ \ 0\ \ 18\ \ 0] \begin{pmatrix} s_{11} & s_{12} & s_{13} & s_{14} \\ s_{21} & s_{22} & s_{23} & s_{24} \\ s_{31} & s_{32} & s_{33} & s_{34} \\ s_{41} & s_{42} & s_{43} & s_{44} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 18 \\ 0 \end{pmatrix}$$

With Equation 3 (which is nothing but Equation 2 written out in full) the linear combination of variances can be obtained. Let us first look on the row vector resulting from `w*ageV`:

$$(s_{11}+s_{31}*18) \quad (s_{12}+s_{32}*18) \quad (s_{13}+s_{33}*18) \quad (s_{14}+s_{34}*18)$$

Now this 4-column vector is again multiplied by `w'`, which result in the expected scalar:

$(s_{11}+s_{31}*18) \quad + \quad (s_{13}+s_{33}*18)*18$ since $s_{13} = s_{31}$

We obtain the same as in the second row of Equation 1:

$s_{11} \quad + \quad 2*s_{31}*18 \quad + \quad s_{33}*18^2$ q.e.d

Study suggestions 2: The reader is invited to try out for interactions between only categorical variables.

Though we will come back to this problem below in section 3.3.1., there is no harm in trying, because the algorithm just presented holds for any kind of weighting, which only depends on the coding of the predictors. We elaborated on this topic for the combination of a categorical and a continuous variable only, because it demonstrates, that there is basically no difference between those types of variables, as long as we know what the numbers used for the coding of a predictor actually mean.

3 Deviation/Effect coding

In some instances we do not want to compare the mean of particular groups with a specified reference group but rather the mean of a subgroup with what is called the *grand mean*. This figure is the mean of all group means, and not the overall mean of the sample observed. Effect coding then enables you to compare the mean of the dependent variable for a given level of a predictor to the mean of the dependent variable for all the levels of the variable. We will focus solely on non-orthogonal designs in non-experimental research.

3.1 One dichotomous predictor

First we will present a simple example with one dichotomous predictor only. We again employ the dichotomous predictor `sktd` (problem definition). To estimate and test the deviation from the *grand mean*, we need a new predictor variable now coded -1 and +1 for either of the 2 categories. We arbitrarily chose +1 for a new variable: `sktdeffect` if `sktd` indicates no definition of mental illness (code 1) and –1 for the other category. The regression analysis now looks like:

Table 24: **`regress dist1sum sktdeffect`**

```
      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  1,  4974) =   48.49
       Model |  2049.24049     1  2049.24049           Prob > F      =  0.0000
    Residual |  210214.383  4974  42.2626424           R-squared     =  0.0097
-------------+------------------------------           Adj R-squared =  0.0095
       Total |  212263.624  4975   42.666055           Root MSE      =   6.501

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  sktdeffect |   -.677629   .0973137    -6.96   0.000    -.8684067   -.4868513
       _cons |   16.57579   .0973137   170.33   0.000     16.38502    16.76657
------------------------------------------------------------------------------
```

First, let us calculate the grand mean:

`display` (17.253423+15.898165)/2 = **16.575794** which is now – as expected – the constant of the regression. Note, that this is **NOT** the total mean (e.g. the mean of all the 4976 observations). The means of the 2 cells can be estimated straightforward by:

$y_1 = c + b_1(-1) =$ **16.575794** $+ 0.677629 = ?$
$y_2 = c + b_1(1) = 16.575794 - 0.677629 = ?$ compare Table 3.

Alternatively, we obtain the deviations from the *grand mean*:

`display` (17.253423+15.898165)/2 -17.253423 = **-.677629**

`display` (17.253423+15.898165)/2 -15.898165 = **.677629**

Of course, the effect is exactly half of the effect obtained from ordinary dummy coding (compare Table 2), but the test is different: the test for b_1 tells us that the deviation from the grand mean is significant.

3.2 Two dichotomous predictors (both coded as deviation from the grand mean)

In case of two dichotomous predictors, we can estimate the 4 cell means either by regarding only the marginal distribution or adopting a fully saturated model as before, with exactly 4 parameters for the 4 cells. The actual means are tabulated in Table 5 and the four cells are now defined by the following linear combination of regression parameters:

$$\hat{y}_{11} = c + (-1)b1 + (-1)b2$$
$$\hat{y}_{12} = c + (1)b1 + (-1)b2$$
$$\hat{y}_{21} = c + (-1)b1 + (1)b2$$
$$\hat{y}_{22} = c + (1)b1 + (1)b2$$

To make things simple, we can use the `e` option of the pre-command `xi3`. Results of the analysis are presented in Table 25 below.

Table 25: **`xi3:regress dist1sum e.sktd e.q0`**

```
e.sktd            _Isktd_1-2          (naturally coded; _Isktd_1 omitted)
e.q0              _Iq0_1-2            (naturally coded; _Iq0_1 omitted)

      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  2,  4973) =  212.94
       Model |  16744.0457     2  8372.02287           Prob > F      =  0.0000
    Residual |  195519.578  4973  39.3162232           R-squared     =  0.0789
-------------+------------------------------           Adj R-squared =  0.0785
       Total |  212263.624  4975   42.666055           Root MSE      =  6.2703

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _Isktd_2 |   -.507479   .0942719    -5.38   0.000    -.6922935   -.3226644
      _Iq0_2 |  -1.726141   .0892854   -19.33   0.000     -1.90118   -1.551103
       _cons |   16.65194   .0939428   177.26   0.000     16.46777    16.83611
------------------------------------------------------------------------------
```

```
predict dp if e(sample)
estimates store distsume
```

Of course, sum of squares, mean squares and all the other characteristics of the model do not depend on the coding scheme; this also holds for the predicted values which are given below

(Table 26). They are exactly the same as those provided by dummy coding (compare Table 7). We should expect that the mean of the predicted means of the 4 cells should be represented by the constant of the regression analysis above.

`display` (18.885563+15.433281+17.870605+14.418323)/4 = **16.651943**

This can be easily verified by calculating the mean employing the definition for each cell:

```
((c   +   (-1)b1    +    (-1)b2) +
 (c   +    (1)b1    +    (-1)b2) +
 (c   +   (-1)b1    +     (1)b2) +
 (c   +    (1)b1    +     (1)b2))/ 4
```

Since all the terms except c cancel out, we obtain: **4c/4** ……….. q.e.d.

Table 26: **`tab sktd q0,sum(dp)`** **`dp`** = predicted y from Table 25.

```
           |
      sktd |  Schizoph   Depressi |     Total
-----------+----------------------+----------
         1 | 18.885563  15.433281 | 17.253422
           |         .          0 | 1.7238418
           |      1733       1554 |      3287
-----------+----------------------+----------
         2 | 17.870605  14.418323 | 15.898164
           |         0          . | 1.7089851
           |       724        965 |      1689
-----------+----------------------+----------
     Total | 18.586487  15.044462 | 16.793408
           | .46280669  .49350943 | 1.8345671
           |      2457       2519 |      4976
```

Without an interaction effect, the conditional differences between columns are necessarily all the same for each row, and this is (statistically) true if the effect for a product of X1 and X2 turns out to be zero.

y_{11} - y_{12} | x2 == -1 = **-2b1**

y_{21} - y_{22} | x2 == +1 = **-2b1**

Since the deviation from the *grand mean* is modelled, the actual difference between the categories is twice the respective parameter. This holds both for the difference with respect to rows and columns. With interaction the equation becomes just the same as for dummy coding, even though the design matrix is different.

If both X1 and X2 are coded equally (-1 or +1) the design for the interaction is necessarily 1. If X2 and X2 are coded differently the numerical value for the product is consistently –1. The estimated means for each of the 4 cells represent the actual means for the 2x2 table.

$$\hat{y}_{11} = c + (-1)b1 + (-1)b2 + (1)b3$$
$$\hat{y}_{12} = c + (1)b1 + (-1)b2 + (-1)b3$$
$$\hat{y}_{21} = c + (-1)b1 + (1)b2 + (-1)b3$$
$$\hat{y}_{22} = c + (1)b1 + (1)b2 + (1)b3$$

Now the difference of means for one of the predictors conditioned on the categories of the other predictor can not be represented any more by one single parameter, as we have just seen for dummy coding.

$$y_{11} - y_{12} \mid X2 == -1 \quad = \quad \mathbf{-2b1 + 2b3}$$

$$y_{21} - y_{22} \mid X2 == +1 \quad = \quad \mathbf{-2b1 - 2b3}$$

One should not forget that both b1 and b2 are **NOT** the same as before, where we estimated the differences under the assumption that they are equal for both categories of the other variable. The numerical results rather look like:

Table 27. **xi3:regress dist1sum e.sktd*e.q0rec**

```
e.sktd            _Isktd_1-2          (naturally coded; _Isktd_1 omitted)
e.q0rec           _Iq0rec_0-1         (naturally coded; _Iq0rec_0 omitted)

      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  3,  4972) =  151.98
       Model |   17829.863     3  5943.28765           Prob > F      =  0.0000
    Residual |  194433.761  4972  39.1057443           R-squared     =  0.0840
-------------+------------------------------           Adj R-squared =  0.0834
       Total |  212263.624  4975   42.666055           Root MSE      =  6.2535

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _Isktd_2 |  -.5455136   .0942959    -5.79   0.000    -.7303752    -.360652
   _Iq0rec_1 |  -1.562662   .0942959   -16.57   0.000    -1.747524   -1.377801
   _Isk2Xq01 |   .4968793   .0942959     5.27   0.000     .3120177    .6817408
       _cons |   16.59575   .0942959   176.00   0.000     16.41089    16.78061
------------------------------------------------------------------------------
```

estimates store **distsumeint**

A likelihood test shows that the interaction should be taken into account. Again all the result are the same as already reported in section 2.2 (also compare Table 8) .

Table 28: `lrtest dist1sume dist1sumeint`

```
Likelihood-ratio test                                  LR chi2(1)  =     27.71
(Assumption: dist1sume nested in dist1sumeint)         Prob > chi2 =    0.0000
```

This likelihood test necessarily yields the same results as with dummy coding of both predictors (Table 10). It should be emphasized that the two different coding schemes represent the same model, as far as prediction and explained variance are concerned. The meaning of the parameters (point of reference) is, of course, different. Statistical inference for the one and only one interaction parameter remains invariant, although the numerical values for these terms are necessarily different.

3.3 Predictors with more than 2 categories

Things become more general if we want to employ a predictor with more than 2 categories. We now take the same concept (`skt`) as before, but with 3 categories. The second category of `sktd` (no definition as mental illness) can be divided into 2 categories, so we obtain:

1. mental illness
2. illness
3. neither 1 nor 2

Let's first look at the table of means and the *grand mean*, which will become of interest in the following:

Table 29: **`tab skt3,sum(dist1sum)`**

skt3	Mean	Std. Dev.	Freq.
1	17.253423	6.4038822	3287
2	15.380424	6.7361216	991
3	16.633238	6.5488024	698
Total	16.793408	6.5319258	4976

To obtain the *Grand Mean* (the mean of the means) you have to type:

```
display  (17.253423 + 15.380424 + 16.633238)/3    =   16.42236
```

3.3.1 Manual Coding

As you see in the example below (Table 30), the regression scheme for effect coding is accomplished by assigning 1 to level 2 for the first comparison (because level 2 is the level to be compared to all), level 1 to level 3 for the second comparison (because level 3 is to be compared to all). Note that a -1 is assigned to level 1 for all three comparisons (because it is the level that is never compared to the other levels) and all other values are assigned a 0. This regression coding scheme yields the comparisons described above. It also means that the effect of the omitted category is simply the negative sum of the other effects. We call the new indicator variables `X2` and `X3`, since the category 1 is adopted for to be the omitted category.

Table 30: Contrast Matrix for 3 categories (first category as reference category)

	New variable 1 (`X2`)	New variable 2 (`X3`)
`skt3`	Level 2 v. Mean	Level 3 v. Mean
1	-1	-1
2	1	0
3	0	1

Based on this scheme, we illustrate how to create `X2` and `X3` and enter these new variables into the regression model.

```
generate x2 = -1 if skt3==1
replace  x2 =  1 if skt3==2
replace  x2 =  0 if skt3==3

generate x3 = -1 if skt3==1
replace  x3 =  1 if skt3==3
replace  x3 =  0 if skt3==2
```

Table 31. **`regress dist1sum x2 x3`**

```
      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  2,  4973) =   31.94
       Model |  2692.03569     2  1346.01785           Prob > F      =  0.0000
    Residual |  209571.588  4973  42.1418838           R-squared     =  0.0127
-------------+------------------------------           Adj R-squared =  0.0123
       Total |  212263.624  4975   42.666055           Root MSE      =  6.4917

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x2 |  -1.041938   .1644163    -6.34   0.000    -1.364266   -.7196092
          x3 |   .2108764    .181612     1.16   0.246    -.1451631     .566916
       _cons |   16.42236   .1133924   144.83   0.000     16.20006    16.64466
```

To calculate the means of the dependent variable shown in Table 29, you have to use the equation for the linear regression model and the obtained coefficients from the last STATA output.

The equation for this model now appears as follows:

$\hat{y}$= cons + b_1(X2) + b_2(X3)

As before, the constant is the *grand mean*. Do not confuse "*reference category*" with "*omitted category*"!!! The mean for the second category is:

$\hat{y}_2$=16.42+-1.04 (1)+0.21(0)
$\hat{y}_2$=15.38 and for the third category :

$\hat{y}_3$=16.42+-1.04(0)+0.21(1)
$\hat{y}_3$=16.63

$\hat{y}_2$ represents the mean of `dist1sum` for people with `skt3=2`. By simply typing: `display 15.38-16.42` we can confirm that this result in a value identical to the regression coefficient of `X2`.

$\hat{y}_3$ represents the mean of dist1sum for people with `skt3=3`. The coefficient for `x3` is the mean for level 3 of `skt3` minus the grand mean, i.e., **16.63 - 16.42 = 0.21**.

You can make your life easier, using the `xi3` pre-command and taking the option `e`, writing `e.skt3` as shown below. With the command `refgroup` you can set the omitted category for the independent variable.

`refgroup skt3 1`

Table 32 **`xi3:reg dist1sum e.skt3`**

```
e.skt3            _Iskt3_1-3          (naturally coded; _Iskt3_1 omitted)

      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  2,  4973) =   31.94
       Model |  2692.03569     2  1346.01785           Prob > F      =  0.0000
    Residual |  209571.588  4973  42.1418838           R-squared     =  0.0127
-------------+------------------------------           Adj R-squared =  0.0123
       Total |  212263.624  4975   42.666055           Root MSE      =  6.4917

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    _Iskt3_2 |  -1.041938   .1644163    -6.34   0.000    -1.364266   -.7196092
    _Iskt3_3 |   .2108764    .181612     1.16   0.246    -.1451631     .566916
       _cons |   16.42236   .1133924   144.83   0.000     16.20006    16.64466
```

As expected, the regression coefficients in Table 32 are identical to those in the example with the manually generated codes (Table 31). Of course, one is often interested in additionally knowing to what extent the omitted category deviates from the grand mean. We know from before that the effect is the negative sum of the other effects, so we can either calculate this "by hand", or type:

`display _b[_cons] - _b[ _Iskt3_2] - _b[ _Iskt3_3]` which will result in: **17.253423**. If we compute the negative sum of the two parameters by the command `display - _b[ _Iskt3_2] - _b[ _Iskt3_3]` we obtain **0.83106117.** The sum of the constant and this negative sum results in :

`display _b[_cons] + _b[ _Iskt3_1] =` **17.253423** (comp. Table 29).

Unfortunately this computation does not provide any statistical test. To obtain the respective t-value, we simply have to change the "omitted category" (using the `refgroup` command) and estimate the coefficients again. If you have no access to the `refgroup` command, you have to type: `char skt3[omit] x` where `x` stands for the category you want to become the omitted one. Verify that everything else remains the same and try without using the nice features of `refgroup` and `char` by creating the necessary variables "by hand".

We also can estimate and test the coefficient for the "omitted category" by means of the command `lincom`, which computes and tests point estimates for linear combination of parameters after estimation.

Table 33: lincom -(_Iskt3_2 + _Iskt3_3)

```
 ( 1) - _Iskt3_2 - _Iskt3_3 = 0

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .8310612   .1308871     6.35   0.000     .5744647    1.087658
------------------------------------------------------------------------------
```

Although we got both the parameter and its standard error by means of the `lincom` command, we will show how to compute the standard error from the Var/Covar matrix of parameters. We employ the procedure already presented in section 2.4. For the example to be found in Table 32 the command `matrix list e(V)` will yield:

Table 34: symmetric e(V)[3,3]

```
             _Iskt3_2     _Iskt3_3        _cons
_Iskt3_2    .02703271
_Iskt3_3   -.02144209    .03298291
   _cons    .00131703    .00726722    .01285784
```

The matrix in Table 34 provides all the information necessary to estimate the standard error of a linear combination of the two parameters. We again have to define a weight vector `w,` to select the adequate variances from the variance/covariance matrix of the parameters. We have to weight down the constant; therefore this row vector is (using STATA mnemonics) defined as:

```
matrix input w =(1 1 0)
```

To estimate the standard error of `_Iskt3_2 + _Iskt3_3,` we have to compute (comp. Equation 2 on page 21)

S_b = `sqrt(w*e(V)*w')` = **0 .13088709**. The actual command looks a bit different, because we need 2 commands. The first one defines the matrix operation which results in a scalar, the second one computes the square root of the one and only element of the matrix `x.`

```
matrix x=w*e(V)*w'
display sqrt(x[1,1])
```

This is exactly the standard error of the test on linear combinations we obtained from the analysis presented in Table 33 .

Study suggestions 3: The reader is encouraged to solve the following matrix equation and to compare with results presented earlier.

$$[1\ 1\ 0] \begin{pmatrix} S_{11} & s_{12} & s_{13} \\ S_{21} & s_{22} & s_{23} \\ S_{31} & s_{32} & s_{33} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

3.3.2 A three-categorical and a dichotomous independent variable (Effect coding and dummy coding combined)

In a 2-dimensional table it is of interest to estimate deviations from a *grand mean* conditioning on another (perhaps dichotomous) characteristic of the sample. This means that we have a combination of effect- and dummy coding. To keep things simple we again adopt `skt3` using deviation/effect coding: `e.skt3` and the dichotomous variable `q0rec`. The reference category for `skt3` as well as for `q0rec` will be set to 1 and 0 respectively. The variables `_Iskt3_2_Iskt3_3` represent the deviation of category 2 or 3 from the grand mean respectively, similar to the example before. The effect for the omitted category is the negative sum of the other parameters. The effect `q0rec` estimates the mean difference between the two case stories (with "schizophrenia" as the reference category. The predicted values of the dependent variable can be obtained by the command: `predict` **`dist1sume`** `if e(sample)`**.** The means of `dist1sum` for each of the 6 cells can be found in Table 11.

Table 35: **`xi3:regress dist1sum q0rec e.skt3`**

```
e.skt3             _Iskt3_1-3          (naturally coded; _Iskt3_1 omitted)

      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  3,  4972) =  142.53
       Model |  16809.1834     3  5603.06113           Prob > F      =  0.0000
    Residual |   195454.44  4972  39.3110298           R-squared     =  0.0792
-------------+------------------------------           Adj R-squared =  0.0786
       Total |  212263.624  4975   42.666055           Root MSE      =  6.2699

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
b1     q0rec |  -3.419012   .1804199   -18.95   0.000    -3.772715    -3.06531
b2  _Iskt3_2 |   -.529245   .1610862    -3.29   0.001     -.845045    -.213445
b3  _Iskt3_3 |  -.1262784   .1763061    -0.72   0.474    -.4719161    .2193594
       _cons |   18.21431   .1446921   125.88   0.000     17.93065    18.49797
```

`predict dist1sume if e(sample)`

Of course, these values are not the actual means of the sample, since the linear model is not saturated. We have 6 cells estimated by only 4 parameters. Assuming that this restricted model is sufficient, we have to answer 3 main questions.

1. What is actually portrayed by the constant `_cons` ?
2. What deviations are actually portrayed by the effects `_Iskt3_2_Iskt3_3` ?
3. What difference is estimated by `q0rec` ?

The answer is straightforward, as for dummy coding. All we need to do is looking on how the 6 cells are assembled by the 4 parameters of the model:

$$
\begin{array}{lllllll|l|}
\cline{8-8}
\hat{y}_{11} & = & c & + & (-1)skt3\_2 & + & (-1)skt3\_3 \;+ & (0)q0rec \\
\hat{y}_{21} & = & c & + & (1)skt3\_2 & + & (0)skt3\_3 \;+ & (0)q0rec \\
\hat{y}_{31} & = & c & + & (0)skt3\_2 & + & (1)skt3\_3 \;+ & (0)q0rec \\
 & & & & & & & \\
\hat{y}_{12} & = & c & + & (-1)skt3\_2 & + & (-1)skt3\_3 \;+ & (1)q0rec \\
\hat{y}_{22} & = & c & + & (1)skt3\_2 & + & (0)skt3\_3 \;+ & (1)q0rec \\
\hat{y}_{32} & = & c & + & (0)skt3\_2 & + & (1)skt3\_3 \;+ & (1)q0rec \\
\cline{8-8}
\end{array}
$$

Table 36: **`tab skt3 q0,sum(dist1sume)`**

```
        Means, Standard Deviations and Frequencies of Fitted values

           | Welcher Fragebogen?
      skt3 |  Schizoph   Depressi |     Total
-----------+----------------------+----------
         1 | 18.869835  15.450822 | 17.253423
           |         .  2.771e-06 | 1.7072295
           |      1733       1554 |      3287
-----------+----------------------+----------
         2 | 17.685066  14.266053 | 15.380424
           |         0          . | 1.6033776
           |       323        668 |       991
-----------+----------------------+----------
         3 | 18.088032   14.66902 | 16.633237
           |         .          . | 1.6916361
           |       401        297 |       698
-----------+----------------------+----------
     Total | 18.586488  15.044462 | 16.793408
           | .45175368     .52847 | 1.8381326
           |      2457       2519 |      4976
```

One can see immediately that the constant is the *grand mean* conditioned that `q0rec` is zero (e.g. the cases story of a schizophrenic episode was provided). The two effects for `skt3` (`_Iskt3_2_Iskt3_3`) now estimate the deviation from this mean, also on the condition that `q0rec` is zero.

There is only one parameter to map the difference(s) between the 2 columns of the 3 x 2 table (Table 36). Therefore the difference of 3.419 must necessarily hold for each row of the table concerning the predicted means. This also means that the category deviations are all equal for both values of `q0rec`. Adopting the parameter notation from Table 35 the decomposition is straightforward:

Table 37: Decomposition table:

skt3		q0rec	
	1	**cons + b2**	**cons + b1+ b2**
	2	**cons + b3**	**cons + b1+ b3**
	3	**cons + (b2+b3)*-1**	**cons + b1 + (b2+b3)*-1**

By comparing the predicted means with the actual means (Table 11 with Table 36), we know, that the restricted model is not sufficient – e.g. the differences produced by two case stories are not the same for each category of `skt3`. Or, to say it the other way round: the deviations of a grand mean are not the same for the two case stories. Again, taking category 1 of `skt3` as the omitted category, we need 2 additional parameters to portray the difference of the difference for category 2 and 3 of `skt3` respectively. Using the `e`-option for `skt3`, we plug in a star between the two predictors to obtain the interaction terms, which yields a saturated model.

Table 38: **`xi3:reg dist1sum q0rec*e.skt3`**

```
e.skt3           _Iskt3_1-3          (naturally coded; _Iskt3_1 omitted)

      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  5,  4970) =   92.50
       Model |  18070.9007     5  3614.18013           Prob > F      =  0.0000
    Residual |  194192.723  4970  39.0729825           R-squared     =  0.0851
-------------+------------------------------           Adj R-squared =  0.0842
       Total |  212263.624  4975   42.666055           Root MSE      =  6.2508

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       q0rec |  -2.652814   .2251294   -11.78   0.000    -3.094167   -2.211461
    _Iskt3_2 |  -.9916186   .2590284    -3.83   0.000    -1.499429   -.4838086
    _Iskt3_3 |  -.4188091    .243418    -1.72   0.085    -.8960158    .0583976
    _Iq0Xsk2 |    .548663    .332421     1.65   0.099    -.1030289    1.200355
    _Iq0Xsk3 |   .9176064   .3563928     2.57   0.010     .2189192    1.616294
       _cons |   17.79038   .1636236   108.73   0.000     17.46961    18.11115
------------------------------------------------------------------------------
```

`estimates store dist1sump`

Now we have 6 parameters for 6 cells, and each of the means is a linear combination which provides all the information we need to interpret the deviations from the grand mean of `skt3` conditioned on the case story. We see that the effect called `q0rec` is the difference between the two vignettes – but conditioned on what value of `skt3` ?

$\hat{y}_{11}$ = c + (-1)skt3_2 + (-1)skt3_3 + (0)q0rec + (0) _Iq0Xsk2 + (0) _Iq0Xsk3
$\hat{y}_{21}$ = c + (1)skt3_2 + (0)skt3_3 + (0)q0rec + (0) _Iq0Xsk2 + (0) _Iq0Xsk3
$\hat{y}_{31}$ = c + (0)skt3_2 + (1)skt3_3 + (0)q0rec + (0) _Iq0Xsk2 + (0) _Iq0Xsk3

$\hat{y}_{12}$ = c + (-1)skt3_2 + (-1)skt3_3 + ***(1)q0rec*** + (0) _Iq0Xsk2 + (0) _Iq0Xsk3
$\hat{y}_{22}$ = c + (1)skt3_2 + (0)skt3_3 + ***(1)q0rec*** + ***(1) _Iq0Xsk2*** + (0) _Iq0Xsk3
$\hat{y}_{32}$ = c + (0)skt3_2 + (1)skt3_3 + ***(1)q0rec*** + (0) _Iq0Xsk2 + ***(1) _Iq0Xsk3***

To calculate the grand mean we type the following:
`display` **(19.20 + 16.80 + 17.37)/3 = 17.79**

which is the grand mean for the left column (`q0rec ==0`) and the grand mean for the condition `q0rec == 1` is

`display` **(15.082 + 14.695 + 15.636)/3 = 15.137**

The difference between the two grand means (one for each column) is--2.652814 which is the effect of `q0rec`. This is different from dummy coding (compare the previous chapter) where – in the case of interaction – the conditional parameter for `q0rec` is valid for the omitted category of `skt3` only. To get all the possible parameters, we recalculate the model with another reference category for `skt3`, and we obtain:

Table 39: **`xi3:reg dist1sum q0rec*e.skt3`**

```
e.skt3            _Iskt3_1-3          (naturally coded; _Iskt3_3 omitted)

      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  5,  4970) =   92.50
       Model |  18070.9007     5  3614.18013           Prob > F      =  0.0000
    Residual |  194192.723  4970  39.0729825           R-squared     =  0.0851
-------------+------------------------------           Adj R-squared =  0.0842
       Total |  212263.624  4975   42.666055           Root MSE      =  6.2508

------------------------------------------------------------------------------
      dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
b1       q0rec |  -2.652814   .2251294   -11.78   0.000    -3.094167   -2.211461
b2    _Iskt3_1 |   1.410428   .1851707     7.62   0.000     1.047411    1.773444
b3    _Iskt3_2 |  -.9916186   .2590284    -3.83   0.000    -1.499429   -.4838086
b4   _Iq0Xsk1 |  -1.466269   .2580308    -5.68   0.000    -1.972124   -.9604152
b5   _Iq0Xsk2 |    .548663    .332421     1.65   0.099    -.1030289    1.200355
         _cons |   17.79038   .1636236   108.73   0.000     17.46961    18.11115
------------------------------------------------------------------------------
```

`estimates store dist1sumpint`

We see that both the constant and the effect for q0rec remains the same, and we will get the deviation of category 1 from the grand mean (and the respective statistical inference). Finally, we will provide a summary table, facilitating the interpretation of both the conditional effects and the interaction terms. A likelihood ratio test shows (Table 41) that the employment of interaction effects is of statistical importance. This test yields a simultaneous judge for the statistical necessity of the two extra parameters. Again, this test is performed using the information for both models stored under the arbitrary user defined names dist1sume and dist1sumpint or dist1sump, since the two models with interaction terms only differ with respect to the omitted category.

Table 40: **`estimates table dist1sume dist1sumpint,star stats (N ll r2) b(%10.3f)`**

```
---------------------------------------------
    Variable |   dist1sume     dist1sumpint   
-------------+-------------------------------
       q0rec |     -3.419***       -2.653***  
    _Iskt3_2 |     -0.529**        -0.992***  
    _Iskt3_3 |     -0.126          -0.419     
    _Iq0Xsk2 |                      0.549     
    _Iq0Xsk3 |                      0.918*    
       _cons |     18.214***       17.790***  
-------------+-------------------------------
           N |   4976.000        4976.000     
          ll | -16193.342      -16177.229     
          r2 |      0.079           0.085     
---------------------------------------------
      legend: * p<0.05; ** p<0.01; *** p<0.001
```

Table 41: **`lrtest dist1sume dist1sumpint`**

```
Likelihood-ratio test                                 LR chi2(2)  =     32.23
(Assumption: dist1sume nested in dist1sumpint)        Prob > chi2 =    0.0000
```

Study suggestions 4: It is left to the reader to estimate both the simple slopes and standard errors for the first category of skt3 defining the appropriate selection vector for the algorithm of Equation 2 on page 21.

The decompositions of each mean again provide the necessary information on interpreting these parameters. The "decomposition table" in Table 42 now looks a bit different. Parameters in italics represent the contribution of the interaction. Parameter notation is adopted from Table 39; cell means can be found in Table 11.

Table 42: Decomposition table (compare Table 37)

skt3		q0rec	
	1	**cons + b2**	**cons + b1+ b2 + *b4***
	2	**cons + b3**	**cons + b1+ b3 + *b5***
	3	**cons + (b2+b3)*-1**	**cons + b1 + (b2+b3)*-1 + *(b4+b5)*-1***

3.3.3 An example from the ESEMED study (6 countries and `dsm_mde` / `mcs12` and `age`)

In this section we will show an example a bit more complicated than those provided above. In the ESEMeD study (comp. 1.2) we have 6 countries which might be of interest to compare with respect to the outcome of the 2 SF12 scores : mental and physical health. For simplicity we will focus on the mental health score as a dependent variable only. For demonstrative purposes we will again present two models: one with two categorical predictors and one with a categorical and a continuous predictor. For a second discrete predictor we employ `dsm_mde` mapping lifetime depression as a dichotomy. The variable `dsm_mde` is already coded 0/1, so we have nothing to do. For a continuous predictor we adopt the age of our respondents called `age_main`. Centering will be discussed in the following.

3.3.3.1 Two categorical predictor (`country` and `dsm_mde`)

The two models are estimated by means of the following commands:

```
xi3:regress mcs12 e.countryn dsm_mde
estimates store mcs12
```

and

```
xi3:regress mcs12 e.countryn*dsm_mde
estimates store mcs12int
```

We can tabulate and compare the results from Table 43 quite easily. We do not present the whole output but rather present both models by the `table` option of the `estimates` command. The first column shows results without and, the second column results for the saturated model. The likelihood ratio test tells that the interaction between the countries and the depression variable is statistically important. If no interaction is defined the differences between depression/no depression should be –6.9643 in each of the countries. Of course this is only true if the restricted model (without interaction) holds, which is not the case, as the `lrtest` clearly shows. We know that the difference between depression/no depression is not the same for each country. These differences are represented by the interaction effects. Since the row variable "country" is coded by deviations from the grand mean, the conditional effect called "`dsm_mde`" now represents the difference of the two grand means. This applies for with and without interaction (which is not true for dummy coding !!), although the exact figure must be different, because the estimated cell means for the restricted model can not be the same as for the saturated model - unless all the interaction effects would be zero.

Interaction effects show that only for Italy and Spain (4 and 6) is the difference in mental health QoL between depressive and non depressive people significantly larger than the average difference. The interaction effects do not test the difference in QoL for depression versus non-depression for a particular country compared to the omitted category, although the output tempts to interpret the table this way (Table 43). The output does not provide a parameter for the omitted category. If one wants to get the deviation from the grand mean, one simply should estimate the model a second time, and adopt another category to be "omitted". From Table 45 we depict that for both the conditional effects and the interaction terms the parameter for the omitted country (Belgium) is the negative sum of the other parameters. In Table 43 and Table 45 the indicator variables are labelled with the initial characters of each country.

Table 43: `estimates table mcs12 mcs12int,stats(N ll  r2) b(%10.3f) star`

```
--------------------------------------------
      Variable |     mcs12        mcs12int
-------------+------------------------------
F _Icountryn_2 |   -0.728***      -0.816***
G _Icountryn_3 |    0.331**        0.343**
I _Icountryn_4 |   -0.865***      -0.735***
N _Icountryn_5 |    0.844***       0.735***
S _Icountryn_6 |   -0.509***      -0.346**
       dsm_mde |   -6.964***      -6.855***
F*D   _Ico2Xds |                   0.469
G*D   _Ico3Xds |                   0.167
I*D   _Ico4Xds |                  -1.054**
N*D   _Ico5Xds |                   0.647
S*D   _Ico6Xds |                  -1.120***
         _cons |   54.839***      54.798***
-------------+------------------------------
             N |  21397.000      21397.000
            ll | -73964.586     -73950.435
            r2 |      0.094          0.095
--------------------------------------------
      legend: * p<0.05; ** p<0.01; *** p<0.001
```

Table 44: `lrtest mcs12 mcs12int`

```
Likelihood-ratio test                 LR chi2(5)  =     28.30
(Assumption: mm nested in mpint)      Prob > chi2 =    0.0000
```

We see that statistics stay the same for both estimations (Table 45), but we now know, that also for Belgium (category 1 of `countryn`) the difference for depression deviates from the grand mean difference significantly. But here the difference is *smaller* than the grand mean, whereas for Italy and Spain the difference is, as said before, *bigger* than the grand mean, which is -6.8554576. The difference between the two categories of `dsm_mde` for Italy is –7.909951, which is –1.0545 bigger than the grand mean. This is exactly the effect for `_Ico4Xds` (comp. Table 45). However, the real differences resulting from the interaction effects are small, only for Italy and Spain are these terms significant, meaning the differences produced by `dsm_mde` are greatest for these countries. All the other effects never exceed the value of 1 on a scale from 9 to 72. Predicted values for both models are provided in Table 46. We also see from Table 45 that the conditional effect for `dsm_mde` does not change, even if the reference category for country is changed. This, of course, is one of the main advantages of effect coding as provided by the option `e` of `xi3`. Information contained in the first column is also provided in the second column of Table 43, so this table is a bit redundant.

Table 45: **`xi3:regress mcs12 e.countryn*dsm_mde (Belgium or Spain omitted)`**
`estimates table mcs12int mcs12intref1,star stats(N ll r2)`

Variable	Omitted 1	Omitted 6
B _Icountryn_1		.82012292***
F _Icountryn_2	-.81566565***	-.81566565***
G _Icountryn_3	.34257119**	.34257119**
I _Icountryn_4	-.73533133***	-.73533133***
N _Icountryn_5	.73455273***	.73455273***
S _Icountryn_6	-.34624986**	
dsm_mde	-6.8554576***	-6.8554576***
B*D _Ico1Xds		.8908707*
F*D _Ico2Xds	.4690798	.4690798
G*D _Ico3Xds	.16708308	.16708308
I*D _Ico4Xds	-1.0544945**	-1.0544945**
N*D _Ico5Xds	.64718754	.64718754
S*D _Ico6Xds	-1.1197266***	
_cons	54.79821***	54.79821***
N	21397	21397
ll	-73950.435	-73950.435
r2	.09538222	.09538222

legend: * p<0.05; ** p<0.01; *** p<0.001

Of course; one can also interpret this table of effects the other way around, saying that the deviations of the countries from the grand mean are not equal for a depressive and a non-depressive subpopulation. But this will certainly not be an interesting way to talk about the results, even though the statistical model for both interpretations is absolutely the same. There is no way to decide empirically which interpretation is a “correct” one, since this model does not distinguish between the predictors in a sense of a “causal” order. Predictors may be addressed to as “factors”; which means that the country “generates” differences regarding the effect of depression; or that depression “generates” differences with respect to the countries. However, this is not a question these models answer.

This holds, of course, for all the other examples in this textbook. Questions on causality can never be answered from the relation between variables, whether or not interactions between the predictors are part of the model. Attempts to solve this problem can be found quite frequently in the literature, but go far beyond the scope of this small booklet.

Table 46: **`tab countryn dsm_mde,sum( mcs12p )`**

countryn	dsm_mde 0	1	Total
1	55.766422	48.802135	54.734783
	.00001208	.	2.4744479
	2053	357	2410
2	54.110615	47.146324	52.567282
	.	.	2.8929641
	2248	640	2888
3	55.169487	48.205196	54.444981
	8.260e-06	.	2.1265082
	3178	369	3547
4	53.973915	47.009628	53.318749
	.0000159	0	2.033323
	4266	443	4709
5	55.682884	48.718594	54.300009
	9.050e-06	.	2.7787871
	1901	471	2372
6	54.329456	47.36517	53.485493
	.00001662	.	2.2729434
	4808	663	5471
Total	54.664553	47.760298	53.714924
	.68023851	.71105427	2.4745809
	18454	2943	21397

`tab countryn dsm_mde,sum(  mcs12intp)`

countryn	dsm_mde 0	1	Total
1	55.618332	49.65374948	54.734781
	.	0	2.1192483
	2053	357	2410
2	53.982544	47.596169	52.567281
	0	.	2.6528983
	2248	640	2888
3	55.140781	48.452408	54.444979
	.	0	2.0422584
	3178	369	3547
4	54.062878	46.152927	53.318748
	.00001426	.	2.3094229
	4266	443	4709
5	55.532764	49.324493	54.300009
	8.010e-06	0	2.4771315
	1901	471	2372
6	54.451962	46.476776	53.485493
	.00001773	0	2.6028717
	4808	663	5471
Total	54.664552	47.760299	53.714923
	.60831955	1.2856862	2.4902594
	18454	2943	21397

3.3.3.2 One categorical and one continuous predictor

We already presented a combination of categorical and continuous predictors in section 2.4 but for the case of dummy coding. Centring of the continuous predictor in order to get a sound interpretation of the conditional effects for the other variable has been briefly addressed, but will become most important in this chapter. We will restrict ourselves to only one example, and again show how not only the parameters of the omitted category can easily be computed as linear combinations of the coefficients at hand, but also how the standard errors can be calculated from the variance/covariance matrix of these parameters, as was outlined the first time in section 3.3.1. This becomes important, since the employment of a continuous variable and interaction terms with other (perhaps categorical) predictors always results in conditional effects only for a particular value (zero) of this continuous predictor. However, it is possible, independent from the centring of the continuous variable, to estimate both the simple slopes and their standard errors for all possible and "impossible" (out of range) values of this variable.

First of all, we should have a look at two models without interaction, one of them with the "original" age (`age_main`) and the second one with mean centred age (`agemean`). Of course, only the constant will be different (Table 47). We show both analyses simultaneously below and will ask why the two constant must be different, and which point these two coefficients actually represent for the data.

Table 47: **`estimates table age_main agemean,star stats(N ll r2)`**

```
--------------------------------------------
    Variable |   age_main         agemean
-------------+------------------------------
 _Icountry_1 |  .91849696***    .91849696***
 _Icountry_2 | -1.2197568***   -1.2197568***
 _Icountry_3 |  .63022879***    .63022879***
 _Icountry_4 | -.46775936***   -.46775936***
 _Icountry_5 |  .47907764**     .47907764**
    age_main |   .0216037***
     agemean |                   .0216037***
       _cons |  52.775337***    53.808473***
-------------+------------------------------
           N |      21397           21397
          ll | -74922.285      -74922.285
          r2 |  .00935887       .00935887
--------------------------------------------
      legend: * p<0.05; ** p<0.01; *** p<0.001
```

We should expect the constant for a regression with `age_main` to be the *grand mean* for an observation with an age of zero. Unfortunately we can not obtain this (predicted) value simply by estimating the predicted values, since there are no respondents with an age of zero. Therefore we must estimate the predicted values "by hand":

```
gen mcs1200=_b[_cons]*_cons +  _b[_Icountry_1]*_Icountry_1 +
_b[_Icountry_2]*_Icountry_2 + _b[_Icountry_3]*_Icountry_3 +
_b[_Icountry_4]*_Icountry_4 + _b[_Icountry_5]*_Icountry_5 +
_b[age_main]*A if e(sample)
```

The `A` is a wildcard which represents a particular value of age, for which we want to estimate the dependent variable. We compute the predicted values twice, once for an age of zero and a second time for the mean age (47.82217). For both computations parameters from the left column of Table 47 are used. Tabulating these predictions with respect to the 6 countries, we obtain in Table 48:

Table 48: tab country,sum(mcs120) for age mean and age zero

Country	Mean	zero	Frequ.
Belgium	54.726971	53.693832	2410
France	52.588715	51.55558	2888
Germany	54.438702	53.405567	3547
Italy	53.340714	52.307579	4709
Netherland	54.287552	53.254414	2372
Spain	53.468185	52.435051	5471
Total	53.714923	52.681788	21397

```
display (54.726971 + 52.588715  + 54.438702 +  53.340714 +
54.287552 +  53.468185)/6    =   53.808473

display (53.693832 + 51.55558 + 53.405567 + 52.307579 +
53.254414 + 52.435051  )/6   =   52.775337
```

Compare these two grand means with the constants in Table 47 ! The difference between the two grand means is nothing but the change in the dependent variable from an age of zero to the mean age of the sample (47.82217 years). If we compute the value for the last term for the mean age, we obtain: `display` 0.0216037*47.82217 = **1.033135** which must be 53.714923 - 52.681788 = …….! Since there is no interaction effect, this value must hold for each country. But we will see immediately that this is not true. Estimating the model with the necessary 5 interaction terms for two different omitted categories, we obtain the following results:

Table 49: **`estimates table belgium spain,star stats(N r2 ll )` without centring of age**

```
---------------------------------------------
    Variable |   Belgium         Spain
-------------+-------------------------------
B _Icountry_1 |                  -.12261529
F _Icountry_2 | -2.2414944***   -2.2414944***
G _Icountry_3 | -1.3045075***   -1.3045075***
I _Icountry_4 |  2.1020507***    2.1020507***
N _Icountry_5 | -.13278772      -.13278772
S _Icountry_6 |  1.6993543***
     age_main |  .03385426***    .03385426***
  B   _Ico1Xag |                  .02160779*
  F   _Ico2Xag |   .0221729**      .0221729**
  G   _Ico3Xag |  .04023132***    .04023132***
  I   _Ico4Xag | -.05454257***   -.05454257***
  N   _Ico5Xag |  .01258843       .01258843
  S   _Ico6Xag | -.04205787***
        _cons |  52.185024***    52.185024***
-------------+-------------------------------
            N |      21397           21397
           r2 |  .01577418       .01577418
           ll | -74852.777      -74852.777
---------------------------------------------
      legend: * p<0.05; ** p<0.01; *** p<0.001
```

A likelihood ratio test tells us, that the $CHI^2(5) = 139.02$. We need not look at a CHI^2 table to believe that this is highly significant. We should briefly recall how it is possible to compute the coefficient and its standard error for France (cat. 1) and for age of 18 using the coefficients form the first column of Table 49. Of course, this could be done easily by estimating the model a second time with a different omitted category for the countries, simultaneously centring the age at 18. However, the missing coefficient can be estimated by computing the negative sum of all the 5 conditional effects for country plus the negative sum of all the 5 interaction effects times 18:

```
display -( _b[_Icountryn_2]+_b[ _Icountryn_3]
        +_b[_Icountryn_4] +_b[_Icountryn_5]
        + _b[_Icountryn_6]) +
     (-(_b[_Ico2Xag]+_b[_Ico3Xag]+ _b[_Ico4Xag]
       +_b[ _Ico5Xag]+ _b[_Ico6Xag])*18) =  .26632494
```

If we also want to get the standard error we define the selection vector **:**

`w = (1 1 1 1 1 0 18 18 18 18 18 0)` and compute the standard error adopting Equation 2 and taking the square root of the resulting scalar. We will obtain **0.29584487** which indeed is want we wanted. Table 50 is the proof of the pudding!

Table 50: **`xi3:regress mcs12 e.country*age18`**

```
      Source |       SS       df       MS              Number of obs =   21397
-------------+------------------------------           F( 11, 21385) =   31.16
       Model |  21943.2666    11  1994.84242           Prob > F      =  0.0000
    Residual |  1369144.14 21385  64.0235747           R-squared     =  0.0158
-------------+------------------------------           Adj R-squared =  0.0153
       Total |  1391087.41 21396  65.0162372           Root MSE      =  8.0015

------------------------------------------------------------------------------
       mcs12 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_Icountryn_1 |   .2663249   .2958449     0.90   0.368    -.3135532     .846203
_Icountryn_2 |  -1.842382   .2726636    -6.76   0.000    -2.376823   -1.307941
_Icountryn_3 |  -.5803438    .257397    -2.25   0.024    -1.084861   -.0758263
_Icountryn_4 |   1.120284    .220636     5.08   0.000     .6878214    1.552748
_Icountryn_5 |    .093804   .3062369     0.31   0.759    -.5064432    .6940512
       age18 |   .0338543   .0034243     9.89   0.000     .0271423    .0405662
    _Ico1Xag |   .0216078   .0085591     2.52   0.012     .0048314    .0383842
    _Ico2Xag |   .0221729   .0081758     2.71   0.007     .0061478     .038198
    _Ico3Xag |   .0402313   .0075068     5.36   0.000     .0255175    .0549451
    _Ico4Xag |  -.0545426   .0065493    -8.33   0.000    -.0673798   -.0417054
    _Ico5Xag |   .0125884     .00888     1.42   0.156     -.004817    .0299939
       _cons |    52.7944   .1172794   450.16   0.000     52.56452    53.02428
------------------------------------------------------------------------------
```

Study suggestions 5: The reader is encouraged to try out different values of age. For instance, for 90 year old respondents ….. you will be surprised !

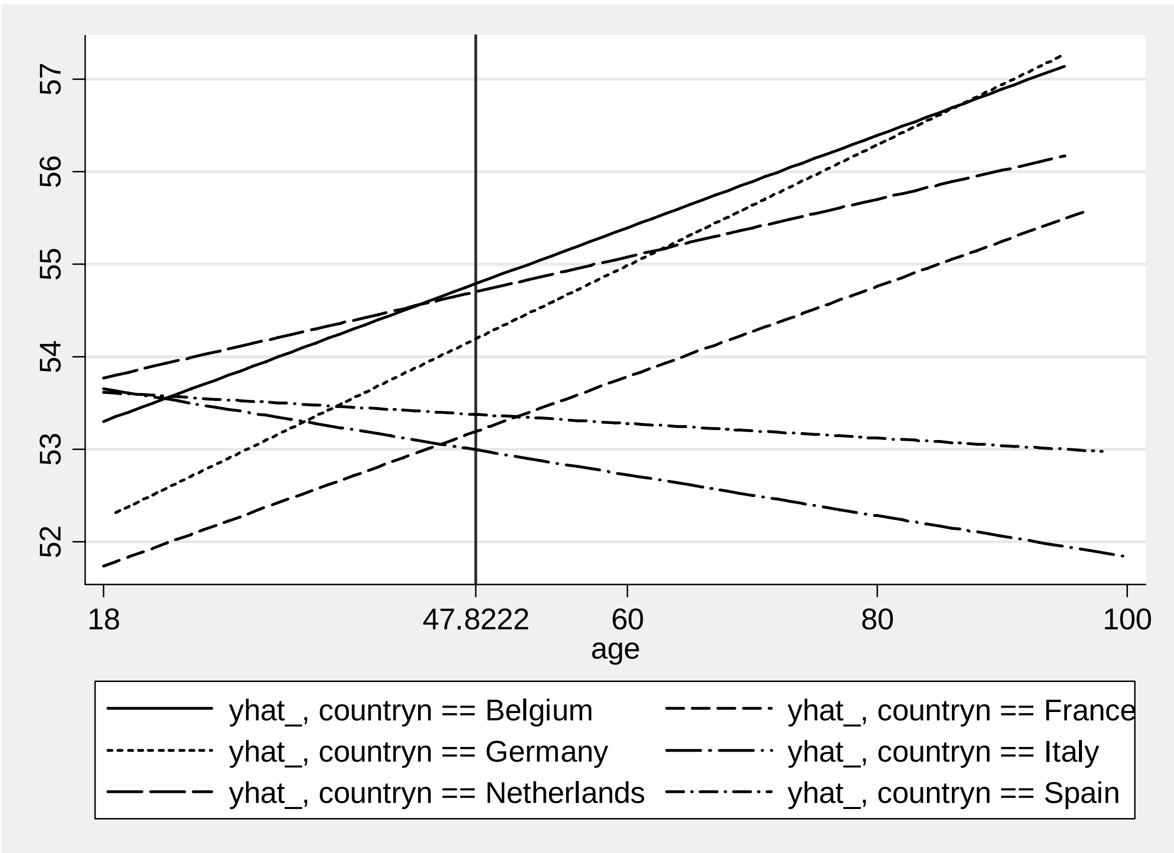
3.3.3.3 Two categorical and one continuous predictor

It might be of interest to further explore the relation between three variables, one of which is continuous. Adding a dichotomous variable "depression" (`dsm_mde`), the saturated model now also contains a three-way interaction. In the first step we will look at a model with Spain as "omitted category" and an additional predictor: `dsm_mde`. Plotting the predicted means will clearly show how the countries differ with respect to the effect of age on mental health related quality of life. In a second step we will estimate and interpret the three-way interaction.

Table 51: **xi3:regress mcs12 e.country*age_main dsm_mde**

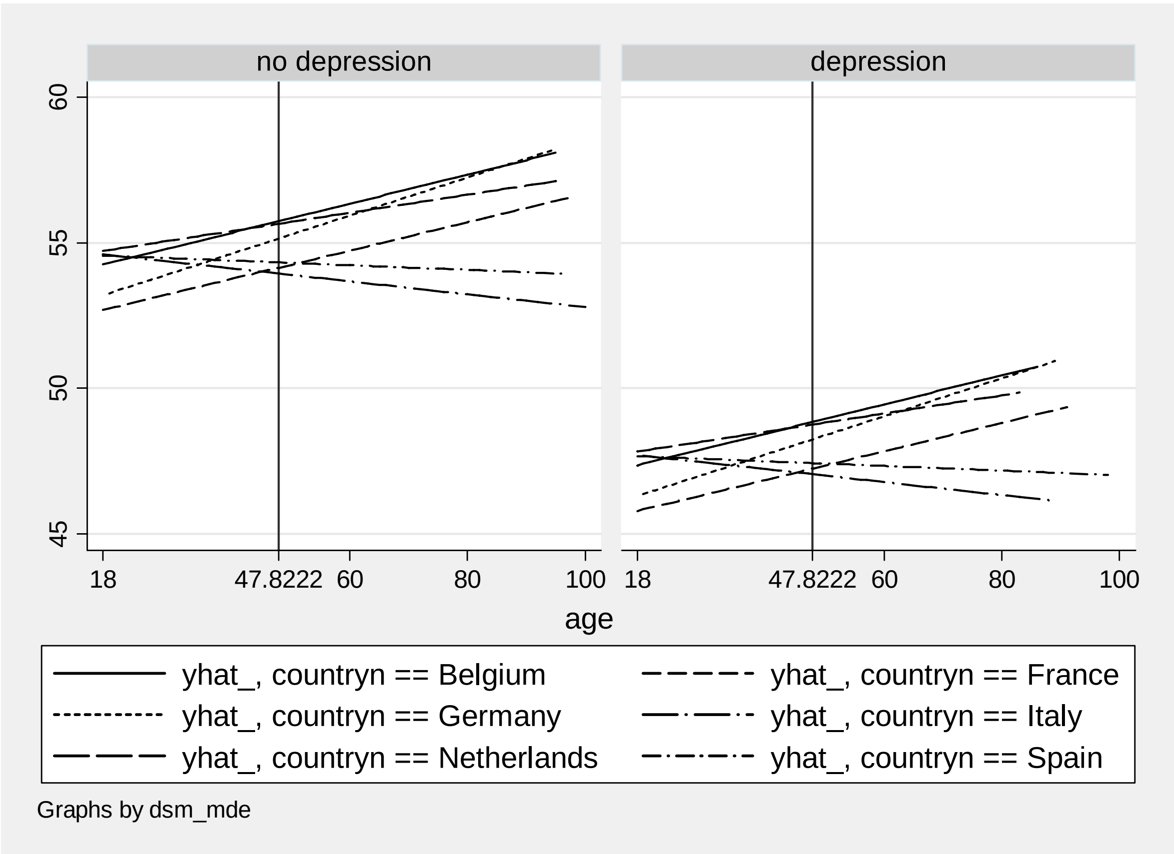
```
e.country         _Icountry_1-6       (naturally coded; _Icountry_6 omitted)

      Source |       SS       df       MS              Number of obs =   21397
-------------+------------------------------           F( 12, 21384) =  199.45
       Model |  140023.914    12  11668.6595           Prob > F      =  0.0000
    Residual |   1251063.5 21384  58.5046529           R-squared     =  0.1007
-------------+------------------------------           Adj R-squared =  0.1002
       Total |  1391087.41 21396  65.0162372           Root MSE      =  7.6488

------------------------------------------------------------------------------
       mcs12 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 _Icountry_1 |  -.1560145   .4174662    -0.37   0.709    -.9742795    .6622506
 _Icountry_2 |  -1.697964   .3894332    -4.36   0.000    -2.461282   -.9346455
 _Icountry_3 |  -1.481953   .3646354    -4.06   0.000    -2.196666   -.7672407
 _Icountry_4 |   1.490272   .3137112     4.75   0.000     .8753741    2.105169
 _Icountry_5 |   .6505748   .4335755     1.50   0.134    -.1992657    1.500415
    age_main |   .0274432   .0032765     8.38   0.000      .021021    .0338654
   _Ico1Xag |   .0223619   .0081819     2.73   0.006     .0063248    .0383989
   _Ico2Xag |   .0212521   .0078155     2.72   0.007     .0059331     .036571
   _Ico3Xag |   .0376565   .0071761     5.25   0.000     .0235907    .0517223
   _Ico4Xag |  -.0495543   .0062617    -7.91   0.000    -.0618276   -.0372809
   _Ico5Xag |   .0036705   .0084909     0.43   0.666    -.0129724    .0203134
     dsm_mde |  -6.888887   .1533398   -44.93   0.000    -7.189444   -6.588329
       _cons |   53.51147   .1685913   317.40   0.000     53.18102    53.84192
------------------------------------------------------------------------------
```

Figure 2: **postgr3 age_main,by(countryn) scheme(s2mono) xlab(18 47.82217 60 80 100) xline(47.82217) two- way interaction**



If we look at the partial plot of the predicted values of `mcs12` against the age, disregarding the predictor `dsm_mde`, we see from Figure 2 that for Italy and Spain the effect of age on `mcs12` is considerably different. For these two countries the effect of age is negative, and it is left for the reader to compute both the conditional effect and the interaction term for Spain, which served as an omitted category. Of course, this prediction does not show the mean differences for depressive and non-depressive respondents, although this predictor is part of the model. The next Figure (Figure 3) shows the two-way interactions for both groups with and without depression, a graphical representation of the model from Table 51 or from Table 52. The particular centring of age is of no importance for both the model fit and the plot.

Figure 3: **`postgr3 age_main, by(country) sep(dsm_mde)` two-way interaction only**



The conditional effects for the countries (the first 5 parameters of Table 51) are meaningless, since they estimate and test these conditional differences for respondents not even born. One will ask quite legitimately why the effect for Italy is positive and for Germany negative, although Figure 2 suspects the difference to be just the other way around. As said before, the conditional effects for the countries test the deviation from the grand mean for the age of zero (!). We already computed the result for an age of 18 in Table 50, and from Figure 2 we can see that these conditional coefficients change dramatically if only the respondents approach

90. Table 52 shows the results for the regression if the age is centred at 90. Also compare results provided in Table 55.

Table 52: **xi3:regress mcs12 e.country*age90 dsm_mde**

mcs12	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_Icountry_1	1.856552	.3695631	5.02	0.000	1.132181	2.580924
_Icountry_2	.2147221	.3599286	0.60	0.551	-.4907648	.9202091
_Icountry_3	1.907129	.3235741	5.89	0.000	1.272899	2.541358
_Icountry_4	-2.969612	.2889784	-10.28	0.000	-3.536031	-2.403192
_Icountry_5	.9809207	.3803656	2.58	0.010	.2353757	1.726466
age90	.0274432	.0032765	8.38	0.000	.021021	.0338654
_Ico1Xag	.0223619	.0081819	2.73	0.006	.0063248	.0383989
_Ico2Xag	.0212521	.0078155	2.72	0.007	.0059331	.036571
_Ico3Xag	.0376565	.0071761	5.25	0.000	.0235907	.0517223
_Ico4Xag	-.0495543	.0062617	-7.91	0.000	-.0618276	-.0372809
_Ico5Xag	.0036705	.0084909	0.43	0.666	-.0129724	.0203134
dsm_mde	-6.888887	.1533398	-44.93	0.000	-7.189444	-6.588329
_cons	55.98136	.1495047	374.45	0.000	55.68832	56.2744

estimates store moint

Table 53: **xi3:regress mcs12 e.country*age_main*dsm_mde**

Source	SS	df	MS
Model	142550.414	23	6197.84408
Residual	1248537	21373	58.4165535
Total	1391087.41	21396	65.0162372

Number of obs	= 21397
F(23, 21373)	= 106.10
Prob > F	= 0.0000
R-squared	= 0.1025
Adj R-squared	= 0.1015
Root MSE	= 7.6431

mcs12	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_Icountry_1	-.0009122	.445809	-0.00	0.998	-.8747313	.8729068
_Icountry_2	-1.693102	.4344548	-3.90	0.000	-2.544666	-.8415378
_Icountry_3	-1.406522	.387742	-3.63	0.000	-2.166526	-.6465189
_Icountry_4	1.493493	.3306763	4.52	0.000	.8453425	2.141643
_Icountry_5	.3571543	.4759232	0.75	0.453	-.5756908	1.289999
age_main	.0279851	.0035087	7.98	0.000	.0211079	.0348624
dsm_mde	-6.697177	.4995717	-13.41	0.000	-7.676375	-5.71798
_Ico1Xag	.0169281	.0086685	1.95	0.051	-.0000628	.0339189
_Ico2Xag	.0193265	.0086491	2.23	0.025	.0023737	.0362794
_Ico3Xag	.0360204	.0075703	4.76	0.000	.0211819	.0508588
_Ico4Xag	-.046617	.0065775	-7.09	0.000	-.0595094	-.0337247
_Ico5Xag	.0072775	.0091802	0.79	0.428	-.0107164	.0252714
_Ico1Xds	-1.723186	1.2987	-1.33	0.185	-4.268736	.8223636
_Ico2Xds	-.0819495	1.002943	-0.08	0.935	-2.047792	1.883893
_Ico3Xds	-.5778693	1.173066	-0.49	0.622	-2.877166	1.721427
_Ico4Xds	.8766869	1.078035	0.81	0.416	-1.236343	2.989717
_Ico5Xds	1.376834	1.19349	1.15	0.249	-.9624951	3.716164
_IagXds	-.0012807	.0102531	-0.12	0.901	-.0213776	.0188163
_Ico1XagXds	.0557987	.0265154	2.10	0.035	.0038265	.1077709
_Ico2XagXds	.0116232	.0206752	0.56	0.574	-.0289018	.0521482
_Ico3XagXds	.0197956	.0245222	0.81	0.420	-.0282698	.067861
_Ico4XagXds	-.0442625	.0220638	-2.01	0.045	-.0875092	-.0010158
_Ico5XagXds	-.0153678	.0247549	-0.62	0.535	-.0638894	.0331537
_cons	53.4427	.1794886	297.75	0.000	53.09088	53.79451

estimates store mo3int

Table 54: **`lrtest moint mo3int`**

```
Likelihood-ratio test                                  LR chi2(12) =   1973.09
(Assumption: moint nested in mo3int)                   Prob > chi2 =    0.0000
```

Now the conditional effect for Italy (`_Icountry_4`) is significantly **below** the average, which can be immediately detected from Figure 2. Whatever centring we choose, further interaction terms tell us that the results shown until now are different for the two subgroups with and without depression. Although the increase in explained variance is only marginal, the LR test is in favour of the model with all possible coefficients. Since a linear effect is adopted for age assuming the proportionality of this effect for each year, this model is not a saturated model ! In Table 53 the coefficients are provided, but without plotting the predicted values for the two groups, as in Figure 3, results are hardly interpretable. Particularly the conditional effects `_Icountry_1 to _Icountry_5` describe the differences between the countries for those without depression (coded zero) and the age of zero, which is of no interest at all.

Figure 4: **`postgr3 age_main,by(countryn) scheme(s2mono) xlab(18 47.82217 60 80 100) xline(47.82217)sep(dsm_mde)`** **three-way interaction**

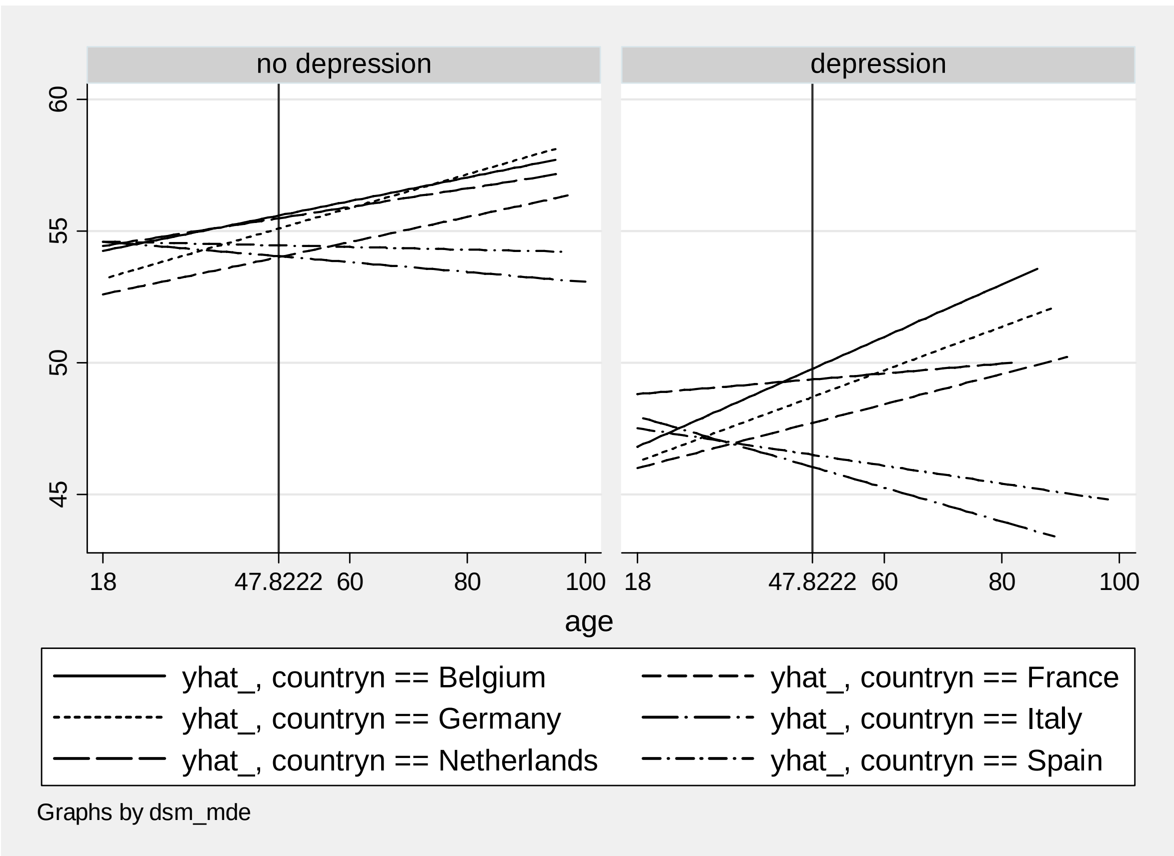


Figure 4 shows that the differences between the 6 countries are much more prominent for those respondents who suffer from life-time depression. For some reasons, Spain and Italy differ even more from all the other nations if the respondents become older and suffer from depression. The biggest difference is observed between respondents from Belgium and Italy for those over 80 years of age. For people younger than 40 the difference between the country seems to be negligible. Only the French (at least those covered by the sample at hand) seem to be an exception.

Results presented above seem to indicate what is called a cohort effect rather than an effect of age "as is". The differences between the different cultures in Europe are smaller for younger people than for those born at the end of the 2nd war or even earlier. Why these differences are even more prominent for those suffering from depression could not be answered here. Even though the LR test is in favour of the 3-way interaction model, we clearly see that there are no essential differences between the two groups indicated by `dsm_mde`. In a last step we should have a look at all the parameters tabulated next to each other, where only the age is centred for different values. Table 55 presents all the necessary information to see which parameters depend on the centring and how this change must be interpreted.

We see immediately that only those coefficients change systematically, which do not contain the age as one term in the product. All the other coefficients obviously depend on the centring of age, because they are conditional effects depending on one particular value (zero) of age. We could ask for the meaning of two further contrasts independent from the centring of age: age*dsm_mde called `_IagXds` in the output. This is the interaction between age and depression for an "average" country, since we employed **effect** coding for the countries. The effect is rather small and we would not see very much from plotting this effect by means of the `postgr3` command. This is an exercise left for the reader.

Table 55: **estimates table mo3int age18 age40 agemean age50 age90 ,star Country: unweighted effect coding; dsm_mde: dummy coded**

Variable	age0	age18	age40	agemean	age50	age90
_Icountry_1	-.00091222	.30379273	.6762099***	.80862403***	.84549042***	1.5226125***
_Icountry_2	-1.6931019***	-1.345224***	-.92003996***	-.76886439***	-.72677448***	.04628744
_Icountry_3	-1.4065224***	-.75815592**	.03429197	.3160494*	.39449556**	1.8353099***
_Icountry_4	1.4934929***	.65438675**	-.37118741**	-.73583363***	-.83735748***	-2.7020378***
_Icountry_5	.3571543	.48814898	.64825358***	.7051793***	.7210284***	1.0121277*
dsm_mde	-6.6971775***	-6.7202292***	-6.7484035***	-6.758421***	-6.76121***	-6.8124361***
_Ico1Xag	.01692805	.01692805	.01692805	.01692805	.01692805	.01692805
_Ico2Xag	.01932655*	.01932655*	.01932655*	.01932655*	.01932655*	.01932655*
_Ico3Xag	.03602036***	.03602036***	.03602036***	.03602036***	.03602036***	.03602036***
_Ico4Xag	-.04661701***	-.04661701***	-.04661701***	-.04661701***	-.04661701***	-.04661701***
_Ico5Xag	.00727748	.00727748	.00727748	.00727748	.00727748	.00727748
_Ico1Xds	-1.7231864	-.7188092	.50876291	.94523016*	1.0667502**	3.2986995**
_Ico2Xds	-.08194949	.12726845	.38297926	.47389808	.49921145	.96414021
_Ico3Xds	-.5778693	-.22154816	.21395544	.36880016	.41191163	1.2037364
_Ico4Xds	.87668694	.07996216	-.89381257*	-1.2400413***	-1.3364374***	-3.106937**
_Ico5Xds	1.3768345	1.1002134	.76212105*	.64191121	.60844269	-.00627074
_IagXds	-.00128065	-.00128065	-.00128065	-.00128065	-.00128065	-.00128065
_Ico1XagXds	.05579873*	.05579873*	.05579873*	.05579873*	.05579873*	.05579873*
_Ico2XagXds	.01162322	.01162322	.01162322	.01162322	.01162322	.01162322
_Ico3XagXds	.01979562	.01979562	.01979562	.01979562	.01979562	.01979562
_Ico4XagXds	-.04426249*	-.04426249*	-.04426249*	-.04426249*	-.04426249*	-.04426249*
_Ico5XagXds	-.01536784	-.01536784	-.01536784	-.01536784	-.01536784	-.01536784
age0	.02798513***					
age18		.02798513***				
age40			.02798513***			
agemean				.02798513***		
age50					.02798513***	
age90						.02798513***
_cons	53.442695***	53.946428***	54.5621***	54.781005***	54.841952***	55.961357***

legend: * p<0.05; ** p<0.01; *** p<0.001

4 Adjacent and Helmert coding

These two forms of contrasts are useful if the predictor can be seen as an ordinal one. Both contrasts are rather similar. The main difference with respect to dummy or effect coding is that the reference category changes for each contrast. This also means that the choice of a reference category is of minor importance, since it does not change the structure of the contrast. Both contrasts exist, as it is called, “forward” and “backward”, but we will focus on “forward” contrasts only, as “backward” coding might by itself explain and the reader is encouraged to estimate this as an exercise. In chapter 6 we will show how these coding schemes fit into a more general scheme of one-way designs in linear models.

4.1 Forward Adjacent coding

The purpose of this coding system is to compare the mean of a criterion for each level (category) of a predictor with the mean regarding the subsequent level (category) of this predictor. For an example we now use the 4 categories of `skt` as outlined in section 1. The allocation of numbers depends on the number of categories, so for a 4 category variable we need to build 3 indicator variables, according to the following scheme:

Table 56: Design matrix for forward adjacent coding

	Categories of `skt`			
	1	**2**	**3**	**4**
A_1 (1 v 2)	.75	-.25	-.25	-.25
A_2 (2 v 3)	.5	.5	-.5	-.5
A_3 (3 v 4)	.25	.25	.25	-.75

Each column of this scheme indicates the weighting coefficients for the linear combination of each cell. The weighting coefficients now are not only zeros or ones (or –1/1) as for dummy and effect coding, but rather a linear combination of proportions which sum up to zero, both for each row and for the row containing the sum total of each column. In section 6.2 (page 67) we will explain the construction of the design matrix in more detail, providing the relevant references. The new variables can be generated by using the prefix `a` `(a.skt)`, again employing the command `xi3`. Tabulating these indicator variables against the predictor of interest will further explain the design.

Table 57: Cross tabulation between the indicator variables for Adjacent coding and `skt`

`tab _Iskt_1 skt`

```
 skt(1 vs. |                      skt
        2) |      1.00       2.00       3.00       4.00 |     Total
-----------+--------------------------------------------+----------
      -.25 |         0      1,816        995        703 |     3,514
       .75 |     1,511          0          0          0 |     1,511
-----------+--------------------------------------------+----------
     Total |     1,511      1,816        995        703 |     5,025
```

`tab _Iskt_2 skt`

```
 skt(2 vs. |                      skt
        3) |      1.00       2.00       3.00       4.00 |     Total
-----------+--------------------------------------------+----------
       -.5 |         0          0        995        703 |     1,698
        .5 |     1,511      1,816          0          0 |     3,327
-----------+--------------------------------------------+----------
     Total |     1,511      1,816        995        703 |     5,025
```

`tab _Iskt_3 skt`

```
 skt(3 vs. |                      skt
        4) |      1.00       2.00       3.00       4.00 |     Total
-----------+--------------------------------------------+----------
      -.75 |         0          0          0        703 |       703
       .25 |     1,511      1,816        995          0 |     4,322
-----------+--------------------------------------------+----------
     Total |     1,511      1,816        995        703 |     5,025
```

The coding scheme follows a simple algorithm, which is outlined in Table 58. The general form of this scheme is presented again for only 4 categories, but can be generalized easily to more then 3 contrasts (rows). The user should keep in mind that the names of the variables generated by `xi3` and one of the prefixes are not different for each contrast; only the labels are which do never show up in the output !

Table 58: Design matrix for forward adjacent coding (k categories)

	1	2	3	4
A_1 (1 v 2)	(k-1)/k	-1/k	-1/k	-1/k
A_2 (2 v 3)	(k-2)/k	(k-2)/k	-2/k	-2/k
A_3 (3 v 4)	(k-3)/k	(k-3)/k	(k-3)/k	-3/k

4.1.1 One 4-categorical predictor

First, lets have look on the means of `dist1sum` for each of the 4 categories of `skt`.

Table 59: **`tab skt,sum(dist1sum)`**

```
            |           Summary of egen
            |     dist1sum=rsum(q15a-q15g) if
            |             distmiss==0
        skt |        Mean   Std. Dev.       Freq.
------------+------------------------------------
       1.00 |   15.788732   6.2545796        1491
       2.00 |   18.469376   6.2721614        1796
       3.00 |   15.380424   6.7361216         991
       4.00 |   16.633238   6.5488024         698
------------+------------------------------------
      Total |   16.793408   6.5319258        4976
```

Table 60: **`xi3:regress dist1sum a.skt`**

```
      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  3,  4972) =   69.53
       Model |  8546.16828     3  2848.72276           Prob > F      =  0.0000
    Residual |  203717.456  4972  40.9729396           R-squared     =  0.0403
-------------+------------------------------           Adj R-squared =  0.0397
       Total |  212263.624  4975   42.666055           Root MSE      =   6.401

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
A_1   _Iskt_1 |  -2.680644   .2242624   -11.95   0.000    -3.120297   -2.240991
A_2   _Iskt_2 |   3.088953   .2532953    12.20   0.000     2.592382    3.585523
A_3   _Iskt_3 |  -1.252814   .3162997    -3.96   0.000    -1.872901   -.6327269
       _cons |   16.56794    .096934   170.92   0.000     16.37791    16.75798
------------------------------------------------------------------------------
```

It is easy to see that the first coefficient represents the difference between category 1 and 2; the second coefficient the difference between 2 and 3, etc. Again, the *grand mean* is represented by the constant.

`display` (15.788732+18.469376+15.380424+16.633238)/4 = **16.567943**

We should now try to compute the (predicted) mean for each of the 4 cells just as we have done before. We can use the syntax of STATA and type:

```
display  _b[_cons]+ _b[ _Iskt_1]*0.75+_b[  _Iskt_2]*0.5 + _b[  _Iskt_3]*0.25
```
15.788732
```
display  _b[_cons]+ _b[ _Iskt_1]*-0.25+_b[  _Iskt_2]*0.5 + _b[  _Iskt_3]*0.25
```
18.469376
```
display  _b[_cons]+ _b[ _Iskt_1]*-0.25+_b[  _Iskt_2]*-0.5 + _b[  _Iskt_3]*0.25
```
15.380424
```
display  _b[_cons]+ _b[ _Iskt_1]*-0.25+_b[  _Iskt_2]*-0.5 + _b[  _Iskt_3]*-0.75
```
16.633238

As expected, we obtain the exact means for the 4 cells (comp. Table 59), since the model has only one predictor and is saturated by default. If there is no access to this feature of STATA, we have to plug in the values for b instead of typing `_b[_cons]+ _b[_Iskt_1] _b[_Iskt_2]` and `_b[_Iskt_3]`.

4.1.2 Two categorical predictors

For this example we again employ the unrestricted variable `skt` just as described in section 1. The actual eight cell means of `dist1sum` can be found in Table 61. The models will contain a mixture of adjacent coded `skt` and dummy coded `q0rec`.

Table 61: **`tab skt q0rec,sum(dist1sum)`**

```
           |        q0rec
       skt |         0          1 |     Total
-----------+----------------------+----------
      1.00 | 18.349177  14.305085 | 15.788732
           | 5.7618022  6.0484204 | 6.2545796
           |       547        944 |      1491
-----------+----------------------+----------
      2.00 | 19.593592  16.283607 | 18.469376
           | 5.9733664  6.2671683 | 6.2721614
           |      1186        610 |      1796
-----------+----------------------+----------
      3.00 | 16.798762  14.694611 | 15.380424
           | 6.9977635  6.5010322 | 6.7361216
           |       323        668 |       991
-----------+----------------------+----------
      4.00 | 17.371571  15.636364 | 16.633238
           |  6.537514  6.4417156 | 6.5488024
           |       401        297 |       698
-----------+----------------------+----------
     Total | 18.586488  15.044462 | 16.793408
           |  6.256263  6.3185563 | 6.5319258
           |      2457       2519 |      4976
```

Two hierarchically nested models using the last category as omitted category were estimated. Predicted means for the first model without interactions is presented in Table 63.

Table 62: **`xi3:regress dist1sum a.skt q0rec`**

```
a.skt             _Iskt_1-4          (naturally coded; _Iskt_4 omitted)

      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  4,  4971) =  123.89
       Model |   19242.248     4    4810.562           Prob > F      =  0.0000
    Residual |  193021.376  4971  38.8294862           R-squared     =  0.0907
-------------+------------------------------           Adj R-squared =  0.0899
       Total |  212263.624  4975   42.666055           Root MSE      =  6.2313

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _Iskt_1 |  -1.780671   .2249509    -7.92   0.000    -2.221674   -1.339668
     _Iskt_2 |   2.063455   .2542043     8.12   0.000     1.565103    2.561808
     _Iskt_3 |  -.4905968   .3113211    -1.58   0.115    -1.100924      .11973
       q0rec |  -3.066468   .1847596   -16.60   0.000    -3.428679   -2.704258
       _cons |   18.15664   .1344144   135.08   0.000     17.89313    18.42015
------------------------------------------------------------------------------
```

```
predict distsuma if e(sample)
(option xb assumed; fitted values)
(49 missing values generated)
```

Table 63: **`tab skt q0rec,sum(dist1suma)`**

```
           |        q0rec
       skt |         0          1 |     Total
-----------+----------------------+----------
      1.00 | 17.730211  14.663744 | 15.788732
           |         .          . | 1.4783799
           |       547        944 |      1491
-----------+----------------------+----------
      2.00 | 19.510883  16.444414 | 18.469377
           |         .          . | 1.4526483
           |      1186        610 |      1796
-----------+----------------------+----------
      3.00 | 17.447428   14.38096 | 15.380424
           |         .          . | 1.4380485
           |       323        668 |       991
-----------+----------------------+----------
      4.00 | 17.938025  14.871556 | 16.633238
           |         .          . | 1.5172068
           |       401        297 |       698
-----------+----------------------+----------
     Total | 18.586488  15.044462 | 16.793408
           | .90290254  .80588189 | 1.9666694
           |      2457       2519 |      4976
```

Adopting the predicted values (`distsuma`) conditioned on `q0rec==0`, we obtain the constant:

`display` (17.730211+19.510883+17.447428+17.938025)/4 = **18.15664**

Estimating the difference between the two grand means, we obtain:

`display` ((14.663744+16.444414+14.38096+14.871556)/4)-
((17.730211+19.510883+17.447428+17.938025)/4) = **-3.0664683**

As in previous chapters the model with and without interaction effects can be compared by storing and tabulating using the `estimates` command. The two models were called `mo1` and `mo2`, without and with interactions, respectively. By employing the `table` feature of `estimates` , we obtain in Table 64:

Table 64: **`estimates table mo1 mo2,star stats(N ll r2 bic aic)`**

```
---------------------------------------------
    Variable |      mo1              mo2
-------------+-------------------------------
b1    _Iskt_1 | -1.7806709***   -1.2444146***  A_1 |
b2    _Iskt_2 |  2.0634552***    2.7948303***  A_2 | q0rec = 0
b3    _Iskt_3 | -.49059681      -.57280946     A_3 |
b4      q0rec | -3.0664683***   -2.7983591***
b5   _Isk1Xq0 |                 -.73410724
b6   _Isk2Xq0 |                 -1.2058345*
b7   _Isk3Xq0 |                  -.3689434
b0      _cons |  18.156637***    18.028275***
-------------+-------------------------------
           N |       4976            4976
          ll | -16162.177       -16151.081
          r2 |  .09065259        .09469899
         bic |  32366.915        32370.261
         aic |  32334.353        32318.161
---------------------------------------------
      legend: * p<0.05; ** p<0.01; *** p<0.001
```

Table 65: lrtest mo1 mo2

```
Likelihood-ratio test                                 LR chi2(3)  =     22.19
(Assumption: mo1 nested in mo2)                       Prob > chi2 =    0.0001
```

Although the likelihood ratio test declares the two models to be statistically different, only one effect turns out to be significant (`_Isk2Xq0`); the change in R^2 is negligible and the BIC also allow you to prefer the more parsimonious model presented in Table 62. However, we should have a closer look at the meaning of both conditional effects and interaction terms in the case of forward adjacent coding of the row defining predictor `skt`. Using the predicted means from the fully saturated model (which are, of course, the actual means of Table 61), we see that the difference between category 2 and 3 is 2.7984 `if q0 == 0`. This difference is 1.588996 `if q0==1` which can be computed from Table 61.

Δ skt_2, skt_3 | q0rec = 1 = b2 + b6

If no interaction effect is employed these difference would be equal (`b2` of `mo1`: 2.0634552); the difference generated by the vignette is **–1.205834** which is represented and tested by the interaction parameter just mentioned before (`b6` of `mo2`).

Study suggestions 6: Verify the meaning of `q0rec` in model `mo2` of Table 64.

4.2 Forward Helmert coding

The idea of the so called Helmert contrast is to compare each level of categorical variable with the ***mean*** of all subsequent levels. This is similar to adjacent coding but makes sense only if we really expect the variable to be ordinal and do not want to inspect whether this assumption really holds. It should be employed with care if we must assume that the different levels of the predictor are ordinal in their own nature but not necessarily with respect to the criterion used as dependent variable. For sake of simplicity we will again use our variable `skt` (with 4 levels). The coding scheme is only of minor difference and looks like:

Table 66: Design matrix for forward Helmert coding (4 categories)

	1	2	3	4
H_1 (1 v 2+)	.75	-.25	-.25	-.25
H_2 (2 v 3+)	.0	0.66	-.33	-.33
H_3 (3 v 4)	0	0	0.5	-0.5

We see that the sum of each row is zero and, furthermore, the sum of the sums of all columns is also zero – again all effects are orthogonal on each other. Tabulating the indicators H_1 to H_3 against the original variable may facilitate the “meaning” of these predictors. It must be mentioned, that the variable names created by `xi3` do not differ, and only the labels tell the user what kind of a contrast was employed.

Table 67: Cross tabulation between the indicator variables for Helmert coding and skt

tab _Iskt_1 skt

```
 skt(1 vs. |                       skt
       2+) |      1.00       2.00       3.00       4.00 |     Total
-----------+--------------------------------------------+----------
      -.25 |         0      1,816        995        703 |     3,514
       .75 |     1,511          0          0          0 |     1,511
-----------+--------------------------------------------+----------
     Total |     1,511      1,816        995        703 |     5,025
```

tab _Iskt_2 skt

```
 skt(2 vs. |                       skt
       3+) |      1.00       2.00       3.00       4.00 |     Total
-----------+--------------------------------------------+----------
 -.3333333 |         0          0        995        703 |     1,698
         0 |     1,511          0          0          0 |     1,511
  .6666667 |         0      1,816          0          0 |     1,816
-----------+--------------------------------------------+----------
     Total |     1,511      1,816        995        703 |     5,025
```

tab _Iskt_3 skt

```
 skt(3 vs. |                       skt
        4) |      1.00       2.00       3.00       4.00 |     Total
-----------+--------------------------------------------+----------
       -.5 |         0          0          0        703 |       703
         0 |     1,511      1,816          0          0 |     3,327
        .5 |         0          0        995          0 |       995
-----------+--------------------------------------------+----------
     Total |     1,511      1,816        995        703 |     5,025
```

The more general coding scheme for k levels of the predictor is shown in Table 68. This example demonstrate the coding scheme again for an example of 4 categories, but can easily be expanded to k categories :

Table 68: Design matrix for forward Helmert coding (k categories)

	1	**2**	**3**	**4**
H_1 (1 v 2+)	(k-1)/k	-1/k	-1/k	-1/k
H_2 (2 v 3+)	.0	((k-1)-1)(k-1)	-1/(k-1)	-1/(k-1)
H_3 (3 v 4)	0	0	((k-2)-1)/(k-2)	-1/(k-2)

4.2.1 One 4- categorical predictor

We employ the same data as in section 4.1.1. The means for each category of `skt` can be found in Table 59. The necessary indicator variables can be obtained by typing `h.[varname]` and using `xi3` just as before.

Table 69: **`xi3:regress dist1sum h.skt`**

```
h.skt             _Iskt_1-4          (naturally coded; _Iskt_4 omitted)

      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  3,  4972) =   69.53
       Model |  8546.16828     3  2848.72276           Prob > F      =  0.0000
    Residual |  203717.456  4972  40.9729396           R-squared     =  0.0403
-------------+------------------------------           Adj R-squared =  0.0397
       Total |  212263.624  4975   42.666055           Root MSE      =   6.401

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _Iskt_1 |  -1.038947   .2028082    -5.12   0.000    -1.436541   -.6413534
     _Iskt_2 |   2.462546   .2186889    11.26   0.000     2.033819    2.891272
     _Iskt_3 |  -1.252814   .3162997    -3.96   0.000    -1.872901   -.6327269
       _cons |   16.56794    .096934   170.92   0.000     16.37791    16.75798
------------------------------------------------------------------------------
```

If we calculate the grand mean of the 4 levels, we obtain:

(15.788732+18.469376+15.380424+16.633238)/4 = **16.567943**

and furthermore:

(18.469376 + 15.380424 + 16.633238)/3 - 15.788732 = **-1.038947** which is just equal to the first parameter (`_Iskt_1`) of the model below.

The reader will certainly believe that the other parameters represent the expected differences. The example presented above is not very exciting, and we should turn immediately to an example with two predictors again; with and without interaction.

4.2.2 Two categorical predictors

We will see that the 2 different coding systems (adjacent and Helmert) provide the same coefficients for the constant and the conditional effect of the dichotomous, dummy coded predictor. Without interactions, the conditional effect of `q0rec` holds for each column; this should not be stressed in the following, since it adds nothing new to what has been said already. On the other hand, it sheds light on the advantage of all the last contrasts over the ordinary dummy coding: Whatever reference (omitted) category is chosen, **conditional** effects keep the same indicating differences of grand means. If one wants to look back again, the cell means of interest are provided in Table 61. The predicted values from this model are the same as for adjacent coding and can be found in Table 63. We estimated both a model with and without interaction, stored them under the names `h1` and `h2` respectively, and tabulate the results next to each other (Table 70). Commands are written as follows:

```
xi3:regress dist1sum h.skt q0rec
predict dist1sumh if e(sample)
estimates store h1
xi3:regress dist1sum h.skt*q0rec
estimates store h2
```

Table 70. **`estimates table h1 h2,stats(r2 N )star`**

```
---------------------------------------------
    Variable |      h1              h2
-------------+-------------------------------
     _Iskt_1 | -.56856636**      .42786913
     _Iskt_2 |  1.8181568***    2.5084256***
     _Iskt_3 | -.49059681      -.57280946
       q0rec | -3.0664683***   -2.7983591***
    _Isk1Xq0 |                  -1.660978***
    _Isk2Xq0 |                 -1.3903062**
    _Isk3Xq0 |                  -.3689434
       _cons |  18.156637***    18.028275***
-------------+-------------------------------
          r2 |  .09065259       .09469899
           N |       4976            4976
---------------------------------------------
      legend: * p<0.05; ** p<0.01; *** p<0.001
```

`lrtest h1 h2`

```
Likelihood-ratio test                                 LR chi2(3)  =     22.19
(Assumption: h1 nested in h2)                         Prob > chi2 =    0.0001
```

Again, the effect for `q0rec` and the constant must differ, since only the model `h2` is a saturated one. The `lrtest` tells that the interaction is statistically necessary. We will find the predicted values from the model h1 (without interaction effect) in Table 71. Computing the first contrast condition for both vignettes, we find, as expected, the same value.

```
display 17.730211 -((19.510883+17.447428+17.938025)/3)        -.56856767

display 14.663744 -((16.444414+14.38096 +14.871556)/3)        -.568566
```

Table 71: `tab skt q0rec,sum(dist1sumh)`

```
        Means, Standard Deviations and Frequencies of Fitted values

           |        q0rec
       skt |         0          1 |     Total
-----------+----------------------+----------
      1.00 | 17.730211  14.663744 | 15.788732
           |         .          . | 1.4783799
           |       547        944 |      1491
-----------+----------------------+----------
      2.00 | 19.510883  16.444414 | 18.469377
           |         .          . | 1.4526483
           |      1186        610 |      1796
-----------+----------------------+----------
      3.00 | 17.447428   14.38096 | 15.380424
           |         .          . | 1.4380485
           |       323        668 |       991
-----------+----------------------+----------
      4.00 | 17.938025  14.871556 | 16.633238
           |         .          . | 1.5172068
           |       401        297 |       698
-----------+----------------------+----------
     Total | 18.586488  15.044462 | 16.793408
           | .90290254  .80588189 | 1.9666694
           |      2457       2519 |      4976
```

For the saturated model things become entirely different. Taking cell means again from Table 61, we now observe two very different contrasts for the values of `q0rec` (Schizophrenia versus depression):

Schizophrenia `display 18.349177-((19.593592+16.798762+17.371571)/3)` **.42786867**

Depression `display 14.305085-((16.283607+14.694611+15.636364)/3)` **-1.233109**

Of course, the difference between these two contrasts is –1.660978, which is nothing but the significant interaction parameter `_Isk1Xq0` from Table 70. Again, the results show that in this case respondents are presented with a schizophrenic episode, those who label the person described as "schizophrenic" show a more pronounced social distance then all the other. For those who had to label the vignette "Depression", things are just the other way round. Therefore, looking only on labelling would only tell us half of the story – it is of great importance **what symptoms** where offered to be evaluated by the respondents.

5 Summary and conclusion

Here we once again have a look at all contrasts next to each other. The second part of Table 72 only confirms that the different contrasts used are all statistically equivalent. This holds, although the model parameters may differ considerably. Modelnames are chosen arbitrarily due to the meaning of the contrast.

Table 72: **estimates table I E A H,star stats(N ll r2 bic aic) b(%10.3f) The capital letter indicates the type of contrast.**

VARIABLE	I	E	A	H
_Iskt_1	0.978*	0.321	-1.244***	0.428
_Iskt_2	2.222***	1.565***	2.795***	2.508***
_Iskt_3	-0.573	-1.230***	-0.573	-0.573
q0rec	-1.735***	-2.798***	-2.798***	-2.798***
_Isk1Xq0	-2.309***	-1.246***	-0.734	-1.661***
_Isk2Xq0	-1.575**	-0.512	-1.206*	-1.390**
_Isk3Xq0	-0.369	0.694	-0.369	-0.369
_cons	17.372***	18.028***	18.028***	18.028***
N	4976.000	4976.000	4976.000	4976.000
ll	-16151.081	-16151.081	-16151.081	-16151.081
r2	0.095	0.095	0.095	0.095
bic	32370.261	32370.261	32370.261	32370.261
aic	32318.161	32318.161	32318.161	32318.161

legend: * p<0.05; ** p<0.01; *** p<0.001

6 User defined coding

In this last chapter two related goals will be pursued. First we want to show how the contrasts, e.g. the contrast- or indicator variables are actually built, and secondly how it is possible to define ones own contrasts, tailored to a particular research question. We decided to deal with this problems at the end of this textbook, simply because the average user never has to carry out all the following procedures manually, but rather employ one of the predefined contrasts using the program adopted for all the analyses above. In any case, it is of great importance to see how the contrast schemes are transformed to design matrices, which afterwards are employed to generate the predictors. We will briefly outline this procedure for two examples already presented above, and one arbitrarily chosen contrast.

To introduce the problem we will briefly summarize the linear one-way model (compare (Finn, 1974: 219pp and 235pp) for a 4- categorical predictor. The model may be written as :

$$y. = A\theta + E$$

We notate y. instead of y because usually more than one person is observed for each level.

$$
y. = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b0 \\ b1 \\ b2 \\ b3 \\ b4 \end{pmatrix} + \begin{pmatrix} e1 \\ e2 \\ e3 \\ e4 \\ e5 \end{pmatrix}
$$

The matrix A is only of rank 4, since the last column is exactly the first column minus the sum of the remaining ones (2,3, and 4). Therefore, only 4 contrasts for θ are uniquely estimable. It is necessary to define linear combinations of parameters by the rows of a so called contrast matrix L. The model matrix A can be defined as the product of the contrast matrix L and the design matrix frequently called K

$$A = KL$$

In order to obtain the matrix of interest K, we have to compute:

$$AL' = KLL'$$

$$AL'(LL')^{-1} = K(LL')(LL')^{-1}$$

$$K = AL'(LL')^{-1}$$

It was shown by Bock (Bock, 1963) that K can be constructed without reference to A. The necessary basis may be calculated from

$K_a = [1, K_{ca}]$ and therefore

$$K_{ca} = L_{ca}'(L_{ca}L_{ca}')^{-1}$$

For sake of simplicity we will let out all subscripts and use L and K respectively. What we have to do is define L , where each row defines one contrast.

6.1 "Dummy coding"

Of course, even ordinary dummy coding can be defined by means of this scheme outlined above. To do this we first have to look at the coding scheme more closely, although it looks so trivial at first glance. The contrast matrix has the form:

Table 73: Contrast matrix `L` for dummy coding

	c1	c2	c3	c4
r1	1	-1	0	0
r2	1	0	-1	0
r3	1	0	0	-1

Writing:

```
matrix input L = (1 -1 0 0 \
                  1 0 -1 0 \
                  1 0  0 -1)
```

and `matrix K= L'*inv(L*L')` we get the necessary information to compute the 3 design variables. Each row portrays the contrast between one of the 3 categories with the reference category (here the category 1 was decided to serve as omitted category). It is not a 0 and 1 matrix anymore, but rather employs a weighting with respect to the number of categories, similar to what is done in the frame of Adjacent – and Helmert coding.

Table 74: Design matrix K for dummy coding

	d1	d2	d3
c1	.25	.25	.25
c2	-.75	.25	.25
c3	.25	-.75	.25
c4	.25	.25	-.75

This matrix provides the necessary information to calculate the three necessary predictors. These predictors (now we call it `d1`, `d2` and `d3`) should estimate and test the differences, the same way as the `i` – option of the `xi3` precommand does. The 3 variables are constructed from the coefficients of each column:

```
gen d1=.25 if skt!=2
gen d2=.25 if skt!=3
gen d3=.25 if skt!=4
replace d1=-.75 if skt==2
replace d2=-.75 if skt==3
replace d3=-.75 if skt==4
```

Table 75: **`regress dist1sum d1 d2 d3`**

```
      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  3,  4972) =   69.53
       Model |  8546.16828     3  2848.72276           Prob > F      =  0.0000
    Residual |  203717.456  4972  40.9729396           R-squared     =  0.0403
-------------+------------------------------           Adj R-squared =  0.0397
       Total |  212263.624  4975   42.666055           Root MSE      =   6.401

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          d1 |  -2.680644   .2242624   -11.95   0.000    -3.120297   -2.240991
          d2 |   .4083086   .2623456     1.56   0.120    -.1060046    .9226217
          d3 |  -.8445054   .2935654    -2.88   0.004    -1.420023   -.2689877
       _cons |   16.56794    .096934   170.92   0.000     16.37791    16.75798
------------------------------------------------------------------------------
```

Table 76: **`xi3:regress dist1sum i.skt`**

```
i.skt             _Iskt_1-4           (naturally coded; _Iskt_1 omitted)

      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  3,  4972) =   69.53
       Model |  8546.16828     3  2848.72276           Prob > F      =  0.0000
    Residual |  203717.456  4972  40.9729396           R-squared     =  0.0403
-------------+------------------------------           Adj R-squared =  0.0397
       Total |  212263.624  4975   42.666055           Root MSE      =   6.401

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _Iskt_2 |   2.680644   .2242624    11.95   0.000     2.240991    3.120297
     _Iskt_3 |  -.4083086   .2623456    -1.56   0.120    -.9226217    .1060046
     _Iskt_4 |   .8445054   .2935654     2.88   0.004     .2689877    1.420023
       _cons |   15.78873   .1657715    95.24   0.000     15.46375    16.11372
------------------------------------------------------------------------------
```

Comparing these results, we surprisingly see, that they are ***not*** identical, particularly not with respect to the constant ! The constant now is not the mean for the omitted category, but rather the grand mean of the four cells. Perhaps this is not what we wanted to obtain but if we estimate a model with several categorical predictors it might advisable to code even simple comparisons like the dummy coding by means of the scheme presented above, since the constant always remains the grand mean. Even though this form of a contrast is hardly ever used, it illustrates what can be done by converting a contrast scheme into a matrix which provides the coefficients for the contrast variables. Nevertheless, `xi3` provides a feature to estimate this contrast by the prefix `g` (Table 77). The constant can be verified by:

`display(15.788732+18.469376+15.380424+16.633238)/4 =` **`16.567943`**

(cell means can be found in Table 59)

Table 77: **xi3:regress dist1sum g.skt**

```
g.skt             _Iskt_1-4         (naturally coded; _Iskt_1 omitted)

      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  3,  4972) =   69.53
       Model |  8546.16828     3  2848.72276           Prob > F      =  0.0000
    Residual |  203717.456  4972  40.9729396           R-squared     =  0.0403
-------------+------------------------------           Adj R-squared =  0.0397
       Total |  212263.624  4975   42.666055           Root MSE      =   6.401

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _Iskt_2 |   2.680644   .2242624    11.95   0.000     2.240991    3.120297
     _Iskt_3 |  -.4083086   .2623456    -1.56   0.120    -.9226217    .1060046
     _Iskt_4 |   .8445054   .2935654     2.88   0.004     .2689877    1.420023
       _cons |   16.56794    .096934   170.92   0.000     16.37791    16.75798
------------------------------------------------------------------------------
```

6.2 Forward Adjacent coding

Lets assume that we do not have a feature like `xi3` in STATA but nevertheless want to model forward adjacent coding as in chapter 4.1. We again use out 4 category variable `skt` and `dist1sum` as criterion. First we have to define a matrix (*contrast matrix*) which represents the contrast coding system. As before, this matrix can then be converted to the design matrix used to compute the variables to be employed as predictor variables. This *contrast matrix* where each *row* represents one contrast appears as follows:

Table 78: **Contrast matrix for forward adjacent coding (each row represents a contrast)**

	c1	c2	c3	c4
r1	1	-1	0	0
r2	0	1	-1	0
r3	0	0	1	-1

Thus, the first line represents the difference between category 1 and 2, the second line the difference between 2 and 3, and the 3rd line the difference between 3 and 4. Each row sums up to zero and the row containing the sum of each column also sums up to zero. This matrix can be defined using the command:

```
matrix input L = (1 -1 0 0 \
                  0 1 -1 0 \
                  0 0 1 -1)
```

In order to obtain the regression coding scheme we need to define the three necessary variables. These are calculated as was done before. Each row yield the coefficients for each contrast variable with regard to each category of a the original predictor. If we list this matrix, we obtain a matrix similar to the matrix provided in Table 56; only rows and columns are interchanged.

Table 79: Design matrix for forward adjacent coding

	a1	a2	a3
c1	.75	.5	.25
c2	-.25	.5	.25
c3	-.25	-.5	.25
c4	-.25	-.5	-.75

By means of these coefficients, we now can define the 3 necessary predictors for the contrast we want to test. Each column of Table 79 provides the coefficients to compute the design variables `a1`, `a2` and `a3`:

```
gen        a1=.75      if skt == 1
replace    a1=-.25     if skt > 1
gen        a2=.5       if skt < 3
replace    a2 =-.5     if skt > 2
gen        a3 =.25     if skt < 4
replace    a3 =-.75    if skt == 4
```

The three variables called `a1`, `a2` and `a3` are equal to those already defined and called A_1 A_2 and A_3. Employing these predictors will provide the same results as what the prefix `a` from `xi3` would have done for the user. Of course, all interpretations hold, too. Compare this result with the results from chapter 4.1.1. Results equal those in Table 60.

Table 80: `regress dist1sum a1 a2 a3`

```
      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  3,  4972) =   69.53
       Model |  8546.16828     3  2848.72276           Prob > F      =  0.0000
    Residual |  203717.456  4972  40.9729396           R-squared     =  0.0403
-------------+------------------------------           Adj R-squared =  0.0397
       Total |  212263.624  4975   42.666055           Root MSE      =   6.401

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          a1 |  -2.680644   .2242624   -11.95   0.000    -3.120297   -2.240991
          a2 |   3.088953   .2532953    12.20   0.000     2.592382    3.585523
          a3 |  -1.252814   .3162997    -3.96   0.000    -1.872901   -.6327269
       _cons |   16.56794    .096934   170.92   0.000     16.37791    16.75798
```

6.3 Forward Helmert Coding

For sake of completeness we present the "manual" coding of the Helmert contrast. The contrast matrix L should be defined as:

```
matrix input L=(1 -.333333 -.333333 -.33333 \
                0 1 -.5 -.5 \
                0 0 1 -1)
```

If we transform this matrix to the design the usual way: `L'*inv(L*L')`, we obtain the matrix K:

Table 81: Design matrix for Helmert coding

	h1	h2	h3
c1	.7500015	7.500e-07	1.125e-06
c2	-.2499995	.66666642	-3.750e-07
c3	-.2499995	-.33333358	.49999963
c4	-.2499995	-.33333358	-.50000037

Table 81 shows the design which is identical to the design matrix in Table 66. Only the rows and columns are interchanged. It seems not to be necessary to compute the empirical example again.

6.4 An arbitrary coding scheme not provided by xi3

6.4.1 One 4-categorical predictor and 3 contrasts

In this last section we will show a contrast which has to be built by oneself, because it is not a feature of the command xi3. Imagine you want to compare the first category of a 4 –category variable with the second category , the first category with the average of category 2 and 3 and, finally, the first category with the mean of all the other categories. The contrast matrix is given below.

Table 82: Contrast matrix L

1	-1	0	0
1	-.5	-.5	0
1	1/3	1/3	1/3

and read in this matrix by means of the STATA command :

```
matrix input L= (1 -1 0 0 \
                  1 -.5 -.5 0 \
                  1 -.33333 -.33333 -.33333)


matrix K= L'*inv(L*L')
```

After transforming `L` to `K` by the same procedure as before, we obtain the coding scheme:

Table 83: Design matrix for one 4 categorical variable and 3 contrasts

```
               u1              u2              u3
c1        -2.221e-16       -.000015          .75001875
c2           -1            -.000015          .75001875
c3            1           -2.000015          .75001875
c4            0            1.999985        -2.2499887
```

By means of these coefficients we create the two variables for the two contrasts the usual way by writing:

```
gen u1=          0        if skt == 1
replace u1=     -1        if skt == 2
replace u1=      1        if skt == 3
replace u1=      0        if skt == 4

gen u2=  -.000015         if skt == 1
replace u2=  -.000015     if skt == 2
replace u2=  -2.000015    if skt == 3
replace u2=   2           if skt == 4
```

```
gen u3 = .75001875         if skt !=4
replace u3= -2.2499887     if skt ==4
```

and from a regression analysis we yield:

Table 84: **`regress dist1sum u1 u2 u3`**

```
      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  3,  4972) =   69.53
       Model |  8546.16828     3  2848.72276           Prob > F      =  0.0000
    Residual |  203717.456  4972  40.9729396           R-squared     =  0.0403
-------------+------------------------------           Adj R-squared =  0.0397
       Total |  212263.624  4975   42.666055           Root MSE      =   6.401

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          u1 |  -2.680644   .2242624   -11.95   0.000    -3.120297   -2.240991
          u2 |  -1.136168    .208614    -5.45   0.000    -1.545143   -.7271922
          u3 |   -1.03895   .2028086    -5.12   0.000    -1.436544   -.6413556
       _cons |   16.56795   .0969344   170.92   0.000     16.37791    16.75798
------------------------------------------------------------------------------
```

The constant is the grand mean, as before. The effect of `u1` should portray the difference between the first and the second category:

`display 15.788732-18.469376 =` **`-2.680644`**

The second contrast `u2` represents the difference between the first category and the mean of second and third category:

`display 15.788732- ((18.469376+15.380424)/2) =` **`1.136168`**

The third contrast estimate the difference between the first category and the means of all other categories.:

`display 15.788732-((18.469376+15.380424+16.633238)/3) =` **`1.0389473`**

However, we could make our life much simpler adopting the `u` option of `xi3`. We could define a pointer by writing:

```
char skt[user] (1 -1 0 0 \ 1 -.5 -.5 0 \ 1 -.33333 -.33333 -.33333)
```

and then estimatinge model again, which yieldsthe same results as before. Unfortunately, the output does not provide any information on how the contrasts are actually defined !! Only the labels of the 3 new variables give the necessary information. One should be careful, because the output does not provide any variable labels, so one never knows – from the output alone – which sort of a contrast was employed.

Table 85: **`xi3:regress dist1sum u.skt`**

```
u.skt             _Iskt_1-4           (naturally coded; _Iskt_4 omitted)

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     _Iskt_1 |  -2.680644   .2242624   -11.95   0.000    -3.120297   -2.240991
     _Iskt_2 |  -1.136168    .208614    -5.45   0.000    -1.545143   -.7271922
     _Iskt_3 |  -1.038944   .2028077    -5.12   0.000    -1.436537   -.6413518
       _cons |   16.56794   .0969344   170.92   0.000     16.37791    16.75798
```

6.4.2 One 4-categorical variable and 3 contrasts and one dichotomous predictor with interaction.

It might be of interest to estimate the contrasts from 6.4.1 dependent on a dichotomous predictor, as we extensively have done before. We use `xi3` to construct the interaction terms and obtain the following output.

Table 86: **`xi3:regress  dist1sum q0rec*u1 q0rec*u2 q0rec*u3`**

```
      Source |       SS       df       MS              Number of obs =    4976
-------------+------------------------------           F(  7,  4968) =   74.24
       Model |  20101.1504     7  2871.59292           Prob > F      =  0.0000
    Residual |  192162.473  4968   38.680047           R-squared     =  0.0947
-------------+------------------------------           Adj R-squared =  0.0934
       Total |  212263.624  4975   42.666055           Root MSE      =  6.2193

------------------------------------------------------------------------------
    dist1sum |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       q0rec |  -2.798346    .195587   -14.31   0.000    -3.181783   -2.414909
          u1 |  -1.244415   .3214451    -3.87   0.000    -1.874589   -.6142403
     _Iq0Xu1 |  -.7341072   .4557537    -1.61   0.107    -1.627586    .1593714
          u2 |   .1530006   .3298555     0.46   0.643    -.4936619     .799663
     _Iq0Xu2 |  -1.337024   .4243911    -3.15   0.002    -2.169018   -.5050305
          u3 |   .4278688   .3136252     1.36   0.173    -.1869751    1.042713
     _Iq0Xu3 |  -1.660981   .4090057    -4.06   0.000    -2.462812   -.8591487
       _cons |   18.02827   .1413196   127.57   0.000     17.75122    18.30532
------------------------------------------------------------------------------
```

`_Iq0Xu1_Iq0Xu2` and `_Iq0Xu1` are nothing but the products of `q0` and one of the `u` terms. If one does not have `xi3`, he or she must compute the three interaction variables “by hand”. Again, we should ask what the constant really represents. Since the dichotomous variable q0 is dummy coded it should be the grand mean conditioned on `q0 == 0`. We calculate this mean using the information from the table in section 4.1.2 .

```
display (18.349177+19.593592+16.798762+17.371571)/4 = 18.028275
```

If we calculate the grandmean for the other row (q0==1), we obtain:

```
display (14.305085+16.283607+14.694611+15.636364)/4 = 15.229917
```

and the difference is –2.798346. This differences is represented by the conditional effect called `q0rec = -2.798346` (First row of Table 86).

6.4.2.1 First contrast u1

The effect `u1` represents the difference between the first category and the second conditioned on `q0==0` (schizophrenia).

```
display (18.349177-19.593592)      -1.244415
```

and conditioned on `q0==1` (depression) this difference is :

```
display (14.305085-16.283607)      -1.978522
```

We see the effect `u1` is not equal for the 2 groups , but rather the difference is:

```
display 1.244415-1.978522               -.734107
```

which is portrayed by the interaction between `u1` and `q0`, called `_Iq0Xu1 = (.7341072)`

Therefore, the interaction term reports the difference between the 2 differences. This interaction is not significant, so we might conclude that the difference between the first and the second category does not depend on `q0` (vignette). In other words, the difference in social distance for those who labelled the person in the cases stories explicitly correct, or only as mentally ill, is significantly different, but this difference is equal for both the schizophrenic episode and the major depression.

6.4.2.2 Second contrast u2

We should briefly have a look at the second contrast u2, which represents the difference between the first category and the average of 1 & 2:

display 18.349177-((19.593592+16.798762)/2) = **0.153**

We see from the table that this difference is not significant as far as the schizophrenia (q0==0) is concerned.

display (14.305085-((16.283607+14.694611)/2)) = **-1.184024**

The difference not only turned out to be the other way round; if the schizophrenia vignette is presented, the "correct diagnosis" exerts a more pronounced (although nonsignificant) social distance. If the case story of a major depression was presented, people show considerably less social distance, if they label the story as depression, compared to the average of the two following categories (mental illness and illness). The interaction effect Iq0Xu2 tells us that the difference is exactly **-1.337024** and highly significant.

Study suggestions 7: Please interpret the results for the third contrast u3.

Of course, we can test whether these 3 interaction terms are simultaneously significant (different from zero):

Table 87: test _Iq0Xu1 _Iq0Xu2 _Iq0Xu3

```
 ( 1)  _Iq0Xu1 = 0
 ( 2)  _Iq0Xu2 = 0
 ( 3)  _Iq0Xu3 = 0

       F(  3,  4968) =    7.40
            Prob > F =    0.0001
```

Comparing the 2 adjusted R^2, we observe an increase of about 5.3%, so we might conclude that the case stories have an effect on the contrasts we wanted to estimate.

7 Reference List

Aiken,L. & West,S.G. (1991). *Multiple Regression: Testing and Interpreting Interaction Effects*. Newbury Park, London, New Delhi: Sage Publications.

Bock,D.R. (1963). Programming Univariate and Multivariate Analysis of Variance. *Technometrics*, **5**, 95-117.

Finn,D.J. (1974). *A General Model for Multivariate Analysis*. New York: Holt, Rinehart and Winston, Inc.

Long,S.J. & Freese,J. (2003). *Regression Models for Categorical Dependent Variable Using Stata*. (2 ed.). College Station ,TX:

Pedhazur,E.J. (1982). *Multiple Regression in Behavioral Research*. (2 ed.). Fort Worth: Hartcourt Brace Jovanovich.

Rabe-Hesketh,S. & Everitt,B.S. (2004). *A Handbook of Statistical Analysis Using STATA*. (3 ed.). Boca Raton, FL: Chapman & Hall /CRC.

Rencher,A.C. (2000). *Linear Models in Statistiscs*. New York: Wiley.

StataCorp. (2005). *Stata Statistical Software: Release 9*. College Station, TX:

8 Appendix

Vignette schizophrenia (`q0rec` = 0)

Imagine that you hear the following about an acquaintance with whom you occasionally spend your leisure time:
Within the past months, your acquaintance appears to have changed. More and more, he retreated from his friends and colleagues, up to the point of avoiding them. If someone managed to involve him in a conversation, he would address only one single topic: the question as to whether some people had the natural gift of reading other people's thoughts. This question became his sole concern. In contrast with his previous habits, he stopped taking care of his appearance and looked increasingly untidy. At work, he seemed absent-minded and frequently made mistakes. As a consequence, he has already been summoned to his boss. Finally, your acquaintance stayed away from work for an entire week without an excuse. Upon his return, he seemed anxious and harassed. He reports that he is now absolutely certain, that people cannot only read other people's thoughts, but that they also directly influence them. He was however unsure who would steer his thoughts. He also said that, when thinking, he was continually interrupted. Frequently, he would even hear those people talk to him, and they would give him instructions. Sometimes, they would also talk to each other and make fun of whatever he was doing at the time. The situation was particularly bad at his apartment, he claimed. At home, he would really feel threatened, and would be terribly scared. Hence he had not spent the night at his place for the past week, but rather he had hidden in hotel rooms and hardly dared to go out.

Vignette major depressive disorder (`q0rec` = 1)

Imagine that you hear the following about an acquaintance with whom you occasionally spend your leisure time:
Within the past two months, your acquaintance has changed in his nature. As opposed to previously, he is down and sad without being able to make out a concrete reason for his feeling low. He appears serious and worried. There is nothing anymore that will make him laugh. He hardly ever talks, and if he says something, he speaks in a low tone of voice about the worries he has with regard to his future. Your acquaintance feels useless and has the impression to do everything wrong. All attempts to cheer him up have failed. He lost all interest in things and is not motivated to do anything. He complains of often waking up in the middle of the night and not being able to get back to sleep. Already in the morning, he feels exhausted and without energy. He says that he encounters difficulty in concentrating on his job. In contrast with previous times, everything takes him very long. He hardly manages his workload. As a consequence, he has already been summoned to his boss.