**Statistical Methods in Epidemiology**

**Lab 3- Solutions.**

**Interactions**

.use "hyper", clear

.desc

.sum

.tab bmi

.tab nomd

Body Mass Index is given by the variable bmi, which takes values 0 for BMI≤30kg/m2 and 1 for >30kg/m2.

The score that indicates the dedication to the Mediterranean diet, is given by the binary variable nomd that takes the value 0: for values ≥5 that indicate high dedication to the Mediterranean diet, and 1 for values <5, that indicate low dedication to the Mediterranean diet. This is in line with our knowledge from theory that variables should be coded as risk factors (Knol et al, 2012).

gen cat\_bmi\_nomd=1 if bmi==0 & nomd==0

replace cat\_bmi\_nomd=2 if bmi==1 & nomd==0

replace cat\_bmi\_nomd=3 if bmi==0 & nomd==1

replace cat\_bmi\_nomd=4 if bmi==1 & nomd==1

1. For estimating interactions in additive form, we first need to fit a logistic regression model:

. logistic hyper i.cat\_bmi\_nomd

Logistic regression Number of obs = 328

LR chi2(3) = 4.63

Prob > chi2 = 0.2008

Log likelihood = -126.97949 Pseudo R2 = 0.0179

------------------------------------------------------------------------------

hyper | Odds Ratio Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

cat\_bmi\_nomd |

2 | 2.610759 1.982514 1.26 0.206 .5893807 11.56479

3 | 1.546875 1.46332 0.46 0.645 .2422322 9.878217

4 | 3.786885 2.977284 1.69 0.090 .8110723 17.68091

|

\_cons | .0606061 .0441345 -3.85 0.000 .014543 .2525674

------------------------------------------------------------------------------

1. Now we can construct the “relative excess risk for interaction” (RERI) index in the additive form

We have ORA=1,B=0= 2.610759

ORA=0,B=1= 1.546875

ORA=1,B=1= 3.786885

For the relative excess risk for interaction we use the following formula:

RERI = ORA=1,B=1 - ORA=1,B=0 - ORA=0,B=1 +1 =

= 3.79 – 2.61 – 1.55 + 1 =0.63>0

In Stata, we use the nlcom command, which assists us when we wish to make nonlinear combinations of estimators:

.nlcom m1\_RERI: exp(\_b[4.cat\_bmi\_nomd])-exp(\_b[2.cat\_bmi\_nomd])-exp(\_b[3.cat\_bmi\_nomd])+1

m1\_RERI: exp(\_b[4.cat\_bmi\_nomd])-exp(\_b[2.cat\_bmi\_nomd])-exp(\_b[3.cat\_bmi\_nomd])+1

------------------------------------------------------------------------------

hyper | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

m1\_RERI | .6292508 1.587599 0.40 0.692 -2.482386 3.740887

------------------------------------------------------------------------------

BMI and Mediterranean diet score have super-additive (since 0.63 is greater than 0) but not at a statistically significant level (since the 95% CI includes 0).

This means that the combined effect (relative excess risk) of low MDScale and high BMI is 0.6 more from the sum of the relative excess risks due to a) low MDScale but low BMI and b)the presence of high BMI but high MDScale (individual effects)

1. For the Synergy index S, the formula is:

S = (RRA=1,B=1-1)/(RRA=1,B=0+ RRA=0,B=1-2)*≈*

*≈* (ΟRA=1,B=1-1)/(ΟRA=1,B=0+ ΟRA=0,B=1-2)

Thus: S =(ΟRA=1,B=1-1)/(ΟRA=1,B=0+ ΟRA=0,B=1-2)=1.29

In Stata:

.nlcom m1\_ln\_SYN: ln(exp(\_b[4.cat\_bmi\_nomd])-1)-ln(exp(\_b[2.cat\_bmi\_nomd])+exp(\_b[3.cat\_bmi\_nomd])-2) , post

m1\_ln\_SYN: ln(exp(\_b[4.cat\_bmi\_nomd])-1)-ln(exp(\_b[2.cat\_bmi\_nomd])+exp(\_b[3.cat\_bmi\_nomd])-2)

------------------------------------------------------------------------------

hyper | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

m1\_ln\_SYN | .2559121 .7671621 0.33 0.739 -1.247698 1.759522

------------------------------------------------------------------------------

For the 95% confidence interval of S, we first estimate the 95% CI for ln(S) following the delta method, i.e. the limits (ln(S1), ln(S2)) and then we exponentiate ln(S1) and ln(S2)

.scalar m1\_Syn\_index\_low95 =exp(\_b[m1\_ln\_SYN]-invnormal(0.975)\*\_se[m1\_ln\_SYN])

.scalar m1\_Syn\_index\_high95=exp(\_b[m1\_ln\_SYN]+invnormal(0.975)\*\_se[m1\_ln\_SYN])

.mat define Model1\_Synergy\_index=(m1\_Syn\_index, m1\_Syn\_index\_low95, m1\_Syn\_index\_high95)

.mat rown Model1\_Synergy\_index= Syn\_index

.mat coln Model1\_Synergy\_index= S\_index S\_low95 S\_high95

.mat list Model1\_Synergy\_index

Model1\_Synergy\_index[1,3]

S\_index S\_low95 S\_high95

Syn\_index 1.2916392 .28716506 5.8096613

In our case, S=1.29, which is greater than 1. This implies super-additive effects. However, the 95% includes 1, which means we cannot reject the null hypothesis of S=1.

This finding means that the combined presence of low MDScale and high BMI results in a 29% increased excess relative risk, as compared to the excess relative risk due to low MDScale but low BMI and the presence of high BMI but high MDScale

1. To get corresponding indices for multiplicative interactions, we need the same logistic regression model.

.logistic hyper i.cat\_bmi\_nomd

Logistic regression Number of obs = 328

LR chi2(3) = 4.63

Prob > chi2 = 0.2008

Log likelihood = -126.97949 Pseudo R2 = 0.0179

------------------------------------------------------------------------------

hyper | Odds Ratio Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

cat\_bmi\_nomd |

2 | 2.610759 1.982514 1.26 0.206 .5893807 11.56479

3 | 1.546875 1.46332 0.46 0.645 .2422322 9.878217

4 | 3.786885 2.977284 1.69 0.090 .8110723 17.68091

|

\_cons | .0606061 .0441345 -3.85 0.000 .014543 .2525674

------------------------------------------------------------------------------

Then we need the following formula:

ΟRA=1,B=1/(ΟRA=1,B=0\*ΟRA=0,B=1)= 3.79 /(2.61\*1.55)

This can be done in Stata, following the steps below:

.nlcom m1\_ln\_mult\_int: \_b[4.cat\_bmi\_nomd]-\_b[2.cat\_bmi\_nomd]-\_b[3.cat\_bmi\_nomd] , post

.scalar m1\_mult\_interaction = exp(\_b[m1\_ln\_mult\_int])

.scalar m1\_mult\_interaction\_low95 = exp(\_b[m1\_ln\_mult\_int]-invnormal(0.975)\*\_se[m1\_ln\_mult\_int])

.scalar m1\_mult\_interaction\_high95 = exp(\_b[m1\_ln\_mult\_int]+invnormal(0.975)\*\_se[m1\_ln\_mult\_int])

.mat define Model1\_mult\_interaction=(m1\_mult\_interaction, m1\_mult\_interaction\_low95, m1\_mult\_interaction\_high95)

.mat rown Model1\_mult\_interaction= mult\_interaction

.mat coln Model1\_mult\_interaction= mult\_interaction m\_int\_low95 m\_int\_high95

. mat list Model1\_mult\_interaction

Model1\_mult\_interaction[1,3]

mult\_inter~n m\_int\_low95 m\_int\_high95

mult\_inter~n .93769167 .12840696 6.8474922

0.94 is smaller than 1, thus we conclude that obesity and low dedication to Mediterranean diet have sub-multiplicative interaction effect on hypertension.

The 95% CI though is (0.13 , 6.85), which includes 1. Thus, we cannot reject the null hypothesis of no multiplicative interaction. In this case, we conclude that the effects do not deviate from multiplicativity.

1. Interactions- Apporach 2
2. We will follow the classic approach (i.e. we will fit a logistic model that includes both bmi, nomd and a term for their interaction) to estimate additive interaction.

.gen bmi\_nomd = bmi\*nomd

.logistic hyper bmi nomd bmi\_nomd

Logistic regression Number of obs = 328

LR chi2(3) = 4.63

Prob > chi2 = 0.2008

Log likelihood = -126.97949 Pseudo R2 = 0.0179

------------------------------------------------------------------------------

hyper | Odds Ratio Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

bmi | 2.610759 1.982514 1.26 0.206 .5893807 11.56479

nomd | 1.546875 1.46332 0.46 0.645 .2422322 9.878217

bmi\_nomd | .9376917 .9512084 -0.06 0.949 .128407 6.847492

\_cons | .0606061 .0441345 -3.85 0.000 .014543 .2525674

------------------------------------------------------------------------------

The formula for RERI in additive form is:

RERI= ORA=1,B=1- ORA=1,B=0- ORA=0,B=1+1

= exp(b1+b2+b3) - exp(b1) - exp(b2) +1 = 0.63

In Stata we do this using, once more, the nlcom command:

. nlcom m2\_RERI: exp(\_b[bmi]+\_b[nomd]+\_b[bmi\_nomd])-exp(\_b[bmi])-exp(\_b[nomd])+1

m2\_RERI: exp(\_b[bmi]+\_b[nomd]+\_b[bmi\_nomd])-exp(\_b[bmi])-exp(\_b[nomd])+1

------------------------------------------------------------------------------

hyper | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

m2\_RERI | .6292508 1.587599 0.40 0.692 -2.482386 3.740887

------------------------------------------------------------------------------

For the synergy index, we have:

S =(ΟRA=1,B=1-1)/(ΟRA=1,B=0+ ΟRA=0,B=1-2)=1.29

. nlcom m2\_ln\_SYN: ln(exp(\_b[bmi]+\_b[nomd]+\_b[bmi\_nomd])-1)-ln(exp(\_b[bmi])+exp(\_b[nomd])-2) ,post

m2\_ln\_SYN: ln(exp(\_b[bmi]+\_b[nomd]+\_b[bmi\_nomd])-1)-ln(exp(\_b[bmi])+exp(\_b[nomd])-2)

------------------------------------------------------------------------------

hyper | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

m2\_ln\_SYN | .2559121 .7671621 0.33 0.739 -1.247698 1.759522

------------------------------------------------------------------------------

. scalar m2\_Syn\_index =exp(\_b[m2\_ln\_SYN])

. scalar m2\_Syn\_index\_low95 =exp(\_b[m2\_ln\_SYN]-invnormal(0.975)\*\_se[m2\_ln\_SYN])

. scalar m2\_Syn\_index\_high95=exp(\_b[m2\_ln\_SYN]+invnormal(0.975)\*\_se[m2\_ln\_SYN])

. mat define Model2\_Synergy\_index=(m2\_Syn\_index, m2\_Syn\_index\_low95, m2\_Syn\_index\_high95)

. mat rown Model2\_Synergy\_index= Syn\_index

. mat coln Model2\_Synergy\_index= S\_index S\_low95 S\_high95

. mat list Model2\_Synergy\_index

Model2\_Synergy\_index[1,3]

S\_index S\_low95 S\_high95

Syn\_index 1.2916392 .28716506 5.8096613

Although we followed a different approach, results are identical to the previously reported (see exercise 1), as expected.

1. Erroneous approach

In a correct approach, variables under study must be coded in such way that they can can be perce4ived as risk factors (i.e. risk of disease increases for greater values of the factor)[Knol et al, 2012].

If we choose to generate a four level variable as we did in exercise I, then (as we actually did), the reference category must correspond to absence of both risk factors, i.e. the lowest risk category.

1. In a wrong approach, we would consider level 2 (bmi>30kg/m2 and Mediterranean diet) as baseline

. char cat\_bmi\_nomd [omit] 2

.

. xi: logistic hyper i.cat\_bmi\_nomd

i.cat\_bmi\_nomd \_Icat\_bmi\_n\_1-4 (naturally coded; \_Icat\_bmi\_n\_2 omitted)

Logistic regression Number of obs = 328

LR chi2(3) = 4.63

Prob > chi2 = 0.2008

Log likelihood = -126.97949 Pseudo R2 = 0.0179

-------------------------------------------------------------------------------

hyper | Odds Ratio Std. Err. z P>|z| [95% Conf. Interval]

--------------+----------------------------------------------------------------

\_Icat\_bmi\_n\_1 | .3830303 .290859 -1.26 0.206 .0864693 1.696696

\_Icat\_bmi\_n\_3 | .5925 .379807 -0.82 0.414 .1686737 2.081275

\_Icat\_bmi\_n\_4 | 1.450492 .5312665 1.02 0.310 .7075362 2.973595

\_cons | .1582278 .0340573 -8.57 0.000 .1037692 .2412666

-------------------------------------------------------------------------------

. nlcom m1\_err\_RERI: exp(\_b[\_Icat\_bmi\_n\_4])-exp(\_b[\_Icat\_bmi\_n\_3])-exp(\_b[\_Icat\_bmi\_n\_1])+1

m1\_err\_RERI: exp(\_b[\_Icat\_bmi\_n\_4])-exp(\_b[\_Icat\_bmi\_n\_3])-exp(\_b[\_Icat\_bmi\_n\_1])+1

------------------------------------------------------------------------------

hyper | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

m1\_err\_RERI | 1.474962 .6332563 2.33 0.020 .2338019 2.716121

------------------------------------------------------------------------------

. nlcom m1\_err\_ln\_mult\_int: \_b[\_Icat\_bmi\_n\_4]-\_b[\_Icat\_bmi\_n\_3]-\_b[\_Icat\_bmi\_n\_1] , post

m1\_err\_ln\_~t: \_b[\_Icat\_bmi\_n\_4]-\_b[\_Icat\_bmi\_n\_3]-\_b[\_Icat\_bmi\_n\_1]

------------------------------------------------------------------------------------

hyper | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------------+----------------------------------------------------------------

m1\_err\_ln\_mult\_int | 1.854948 1.014415 1.83 0.067 -.1332683 3.843165

------------------------------------------------------------------------------------

. scalar m1\_err\_mult\_interaction = exp(\_b[m1\_err\_ln\_mult\_int])

. scalar m1\_err\_mult\_interaction\_low95 = exp(\_b[m1\_err\_ln\_mult\_int]-invnormal(0.975)\*\_se[m1\_err\_ln\_mult\_int])

. scalar m1\_err\_mult\_interaction\_high95 = exp(\_b[m1\_err\_ln\_mult\_int]+invnormal(0.975)\*\_se[m1\_err\_ln\_mult\_int])

. mat define Model1\_err\_mult\_interaction=(m1\_err\_mult\_interaction, m1\_err\_mult\_interaction\_low95, m1\_err\_mult\_interaction\_high95)

. mat rown Model1\_err\_mult\_interaction= mult\_interaction

. mat coln Model1\_err\_mult\_interaction= mult\_interaction m\_int\_low95 m\_int\_high95

. mat list Model1\_err\_mult\_interaction

Model1\_err\_mult\_interaction[1,3]

mult\_inter~n m\_int\_low95 m\_int\_high95

mult\_inter~n 6.3913675 .87523022 46.672953

Then the multiplicative interaction would be equal to 6.4, with 95% confidence interval of (0.88-46.7).

1. Same problems would arise if we followed a classic approach with reference level of nomd indicating no dedication to Mediterranean diet.

. gen err\_nomd=1-nomd

. gen err\_bmi\_nomd= bmi\*err\_nomd

. logistic hyper bmi err\_nomd err\_bmi\_nomd

Logistic regression Number of obs = 328

LR chi2(3) = 4.63

Prob > chi2 = 0.2008

Log likelihood = -126.97949 Pseudo R2 = 0.0179

------------------------------------------------------------------------------

hyper | Odds Ratio Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

bmi | 2.448087 1.64661 1.33 0.183 .6550878 9.148593

err\_nomd | .6464646 .6115454 -0.46 0.645 .1012328 4.12827

err\_bmi\_nomd | 1.066449 1.081821 0.06 0.949 .1460389 7.78774

\_cons | .09375 .0566069 -3.92 0.000 .0287084 .3061491

------------------------------------------------------------------------------

. logit hyper bmi err\_nomd err\_bmi\_nomd

Iteration 0: log likelihood = -129.29544

Iteration 1: log likelihood = -127.08769

Iteration 2: log likelihood = -126.97978

Iteration 3: log likelihood = -126.97949

Iteration 4: log likelihood = -126.97949

Logistic regression Number of obs = 328

LR chi2(3) = 4.63

Prob > chi2 = 0.2008

Log likelihood = -126.97949 Pseudo R2 = 0.0179

------------------------------------------------------------------------------

hyper | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

bmi | .8953071 .6726108 1.33 0.183 -.422986 2.2136

err\_nomd | -.4362368 .9459843 -0.46 0.645 -2.290332 1.417858

err\_bmi\_nomd | .0643341 1.014415 0.06 0.949 -1.923882 2.052551

\_cons | -2.367124 .6038074 -3.92 0.000 -3.550564 -1.183683

------------------------------------------------------------------------------

. logit hyper bmi nomd bmi\_nomd

Logistic regression Number of obs = 328

LR chi2(3) = 4.63

Prob > chi2 = 0.2008

Log likelihood = -126.97949 Pseudo R2 = 0.0179

------------------------------------------------------------------------------

hyper | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

bmi | .9596412 .759363 1.26 0.206 -.5286829 2.447965

nomd | .4362368 .9459843 0.46 0.645 -1.417858 2.290332

bmi\_nomd | -.0643341 1.014415 -0.06 0.949 -2.052551 1.923882

\_cons | -2.80336 .7282191 -3.85 0.000 -4.230644 -1.376077

------------------------------------------------------------------------------

1. Interpretation of interactions is problematic, since they now refer to one risk and one protective factor.