**Statistical Methods in Epidemiology**

**Lab 3.**

**Interactions**

1. Interactions- Apporach 1

In the genetic dataset, data of 328 records of individuals are stored. The outcome of interest is hypertension, (variable hyper). We want to examine whether the development of hypertension is associated with a person’s Body Mass Index, a score indicating the dedication to Mediterranean diet, categorised in two groups and, most importantly, whether and how these two characteristics interact with each other.

1. Open the datafile genetic and inspect it. Check the variables’ coding. Is it meaningful in the context of an interaction exploration?

use "hyper", clear

desc

sum

In order to study the interaction between BMI and dedication to Mediterranean diet, we need to generate a single variable that combines information from both the examined variables. The new variable should be equal to 1 for both factors being absent and 4 for both factors being present.

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. After creating the cat\_bmi\_nomd variable, we can study the interaction following the first method and we are going to estimate interactions in additive form.

logistic hyper i.cat\_bmi\_nomd

1. After fiting the logistic regression model, we can construct the “relative excess risk for interaction” (RERI) index. This can be done by combining the model estimate using the nlcom Stata command for the nonlinear combinations of estimators

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

we can give a name to the combined estimate:

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

1. In a similar way the Synergy (S) index can be estimated.

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

The point estimate, as well as the lower and upper limits of a 95% confidence interval can be assigned to scalars, so that we can use them later in our analysis.

scalar m1\_Syn\_index =exp(\_b[m1\_ln\_SYN])

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

1. Now we will get corresponding indices for multiplicative interactions:

logistic hyper i.cat\_bmi\_nomd

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

1. Interactions- Apporach 2
2. In this exercise, we are going to explore interaction, too. This time 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)

gen bmi\_nomd = bmi\*nomd

logistic hyper bmi nomd bmi\_nomd

1. Using the corresponding formulas for the additive interaction, we get:

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

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

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

1. Erroneous approach
2. In the previous, correct approach, we made sure that both factors under study took values on the same directions, i.e. baseline level indicated in both factors the lower risk category. Let us now see what would have happened if instead of the correct approach we followed so far, we had set as the baseline category of the four-level variable that combines bmi and diet for the second one (i.e. those who have low BMI but have low score for the dedication to the Mediterranean diet)

char cat\_bmi\_nomd [omit] 2

and fit the same logistic regression model

xi: logistic hyper i.cat\_bmi\_nomd

Estimate RERI and Synergy indices, using same formulas as before:

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

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

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

xi: logistic hyper i.cat\_bmi\_nomd

lincom \_b[\_Icat\_bmi\_n\_4]-\_b[\_Icat\_bmi\_n\_3]-\_b[\_Icat\_bmi\_n\_1]

1. Please try a classic logistic regression, now considering a different coding for the mediterranean diet dedication considering a low scroe as the refrnece category (i.e nomd: dedication 1: ≥5, no dedication 0: <5)

gen err\_nomd=1-nomd

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

logistic hyper bmi err\_nomd err\_bmi\_nomd

logit hyper bmi err\_nomd err\_bmi\_nomd

logit hyper bmi nomd bmi\_nomd