**Lab 3 Solutions**

# Part I. Additive Interaction (RERI and Synergy Index)

1.1. Initial Data Exploration

**1.**

. use hyper, clear

. desc

Observations: 328 Diet data with dates

Variables: 4 9 Jan 2015 11:44

--------------------------------------------------------------------------------------Variable Storage Display Value

name type format label Variable label

--------------------------------------------------------------------------------------id float %9.0g Subject identity number

hyper float %9.0g Outcome: 1= hyper, 0= otherwise

nomd float %9.0g risk factor: 1: score<5 , 0: score>=5

bmi float %9.0g risk factor: 1: BMI>30kg/m2 , 0: BMI<=30kg/m2

--------------------------------------------------------------------------------------Sorted by: id

The dataset *hyper* contains 328 observations and 4 variables corresponding to participants’ identifiers, hypertension status, BMI category, and Mediterranean diet adherence.

The variable **nomd** is a binary indicator of Mediterranean Diet Score: 0 = score ≥ 5 (high adherence) and 1 = score < 5 (low adherence).

**2.**

. tab bmi

risk |

factor: 1: |

BMI>30kg/m2 |

, 0: |

BMI<=30kg/m |

2 | Freq. Percent Cum.

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

0 | 70 21.34 21.34

1 | 258 78.66 100.00

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

Total | 328 100.00

. tab nomd

risk |

factor: 1: |

score<5 , |

0: score>=5 | Freq. Percent Cum.

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

0 | 218 66.46 66.46

1 | 110 33.54 100.00

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

Total | 328 100.00

**3.** According to epidemiological convention (Knol et al., 2012), variables representing risk factors should be coded so that higher values correspond to a higher level of risk. This ensures clear interpretation of interaction effects.

1.2. Variable Creation and Coding

**1.** Created a categorical variable combining both risk factors:

. gen cat\_bmi\_nomd = 1 if bmi == 0 & nomd == 0 // Neither risk factor

. replace cat\_bmi\_nomd = 2 if bmi == 1 & nomd == 0 // High BMI only

. replace cat\_bmi\_nomd = 3 if bmi == 0 & nomd == 1 // Low adherence only

. replace cat\_bmi\_nomd = 4 if bmi == 1 & nomd == 1 // Both risk factors

**2.**

. tab bmi nomd

. tab cat\_bmi\_nomd

The cross-tabulation confirms that:

* All four exposure combinations are correctly classified
* No missing or unexpected categories are present
* **Category 1** of cat\_bmi\_nomd represents the **absence of both risk factors** (lowest-risk group) and serves as the **baseline** in the logistic models (i.cat\_bmi\_nomd, base = 1).
* Category 4 corresponds to the presence of both risk factors (highest-risk group).

***Note:*** This coding scheme aligns with epidemiological best practices where:

* Higher categories indicate increased risk
* The reference category represents the absence of both risk factors

1.3. Additive Interaction on the OR Scale (RERI)

**1.**

. 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

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

Note: \_cons estimates baseline odds.

The logistic regression results show increased odds compared to the reference group, with category 4 (both risk factors) showing the highest odds ratio (3.79).

**2. RERI (Relative Excess Risk due to Interaction)**

* **Reminder:** The Relative Excess Risk due to Interaction (RERI) quantifies the departure from additivity of effects on the risk (or rate) ratio scale.
* It measures how much the combined effect of two exposures exceeds or falls short of the sum of their individual effects.
* **Formula:**
  + - * When the outcome is rare, **odds ratios approximate risk ratios**, so:
* **Interpretation:**
* RERI = 0 no additive interaction
* RERI > 0 indicates super-additive interaction → synergy *on the additive scale*
* RERI < 0 indicates sub-additive interaction → antagonism

**Notes**

* Ensure both exposures are **risk-coded** (1 = higher risk).
* The reference category is **00** (neither exposure present).

Visual summary of the joint exposure categories and corresponding odds ratios (OR):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Exposure Combination** | **BMI** | **Mediterranean Diet** | **OR (vs. reference)** | **Symbol** |
| Neither exposure | 0 | 0 (High adherence) | 1.00 (reference) | OR₀₀ |
| High BMI only | 1 | 0 (High adherence) | 2.61 | OR₁₀ |
| Low MedDiet only | 0 | 1 (Low adherence) | 1.55 | OR₀₁ |
| Both exposures | 1 | 1 (Low adherence) | 3.79 | OR₁₁ |

We calculate the Relative Excess Risk due to Interaction (RERI) using the previous logistic regression output:

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

. nlcom m1\_RERI\_OR: 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

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

hyper Coefficient Std. err. z P>z [95% conf. interval]

m1\_RERI .6292508 1.587599 0.40 0.692 -2.482386 3.740887

**4. Interpretation of RERI = 0.63**

* **Direction:** The positive value suggests that the combined effect of high BMI and low Mediterranean diet adherence is *greater* than the sum of their individual effects (possible synergy).
* **Statistical significance:** The 95% CI includes 0 ⇒ no statistically significant additive interaction.
* **Practical meaning:** While not significant, the direction implies a potential synergistic effect worth investigating in larger studies.
* **Clinical implication:** Patients with both risk factors may experience higher risk than expected under additivity.
* **Uncertainty:** The wide CI highlights limited precision, suggesting sample size or power limitations.

1.4. Synergy Index Calculation (S)

* **Reminder:** The Synergy Index (S) quantifies departure from additivity by comparing the excess risk under joint exposure to the sum of excess risks from each exposure separately.
* **Formula:**
* When the outcome is rare, odds ratios can be used instead of risk ratios:
* **Interpretation:**
* S = 1: No additive interaction
* S > 1: Positive (super-additive) interaction (synergy)
* S < 1: Negative (sub-additive) interaction (antagonism)

**Notes**

* For additive interaction measures such as S, both exposures must be coded in the same risk direction (typically 1 = higher hypothesised risk), and the reference category must represent the joint absence of both exposures (00).
* S is strictly ≥ 0 by definition.
* For statistical inference, confidence intervals for **S** should be obtained by applying the delta method to **ln(S)** and then exponentiating back to the original scale.
* When **S** is computed from *odds ratios* (e.g., via logistic regression), it reflects additivity on the OR scale. Interpreting S as a *risk-based* measure requires the rare-disease assumption.

**1.** Based on the formula of synergy and the previous logistic regression output we can calculate that S = 1.29.

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

. 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

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

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

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

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

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

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

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

hyper | Coefficient Std. err. z P>|z| [95% conf. interval]

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

m1\_Syn\_index | 1.291639 .9908967 1.30 0.192 -.6504826 3.233761

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

**3.** 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])

**Notes:**

* **Stata scalars: what and why**

In Stata, the scalar command is used to create and store a numeric value that can be referenced later in your analysis. It is a way to save a single number under a specific name for future use.

* **Why take logs for the Synergy Index (S)?**

For the Synergy Index (S), which is bounded below by 0 and has a typically skewed sampling distribution, we use the natural logarithm transformation (ln) when calculating confidence intervals. This transformation stabilizes variance and ensures that back-transformed confidence intervals remain in the valid parameter space (S > 0). This approach is similar to that used for other ratio measures like odds ratios and relative risks.

* **Delta method for CI of**

The delta method is a statistical technique based on Taylor series expansion that approximates the variance of a function of random variables. In this context, we use it to estimate the variance of ln(S), allowing us to construct confidence intervals on the log scale before back-transforming to the original scale. This method is particularly valuable when working with nonlinear functions of parameter estimates.

**4.** With the following commands, we create and manipulate a matrix in Stata to store and display the Synergy Index results.

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

* *Creates a 1×3 matrix named "Model1\_Synergy\_index"*
* *Populates it with the point estimate and confidence interval bounds previously calculated as scalars*

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

* *Assigns the row name "Syn\_index" to the matrix*

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

* *Assigns column names to the matrix elements:*
* *S\_index: point estimate*
* *S\_low95: lower 95% CI bound*
* *S\_high95: upper 95% CI bound*

. mat list Model1\_Synergy\_index

* Displays the formatted matrix with its row and column labels

Model1\_Synergy\_index[1,3]

S\_index S\_low95 S\_high95

Syn\_index 1.2916392 .28716506 5.8096613

**5. Interpretation:**

**a. Point estimate**

The estimated Synergy Index (S = 1.29) indicates that the joint effect of high BMI and low Mediterranean diet adherence is about 29% larger than expected under additivity.

*Example: If each factor increases risk by 50% independently, under no interaction we would expect their combined effect to be a 100% increase. With S=1.29, the joint effect corresponds to RR₁₁ ≈ 2.29 (129% increase), indicating synergy on the additive scale.*

**b. Statistical significance**

The 95% CI for S is **0.29 to 5.81**, which includes 1 ⇒ we **cannot reject** the null of no additive interaction (α=0.05). The wide CI implies substantial uncertainty.

**c. Clinical relevance**

Despite non-significance, **S>1** is *compatible with* a positive departure from additivity (potential synergy) and may be clinically relevant, but the wide CI argues for **caution** and for **larger, better-powered studies** before drawing firm conclusions.

**Note:** The Synergy Index complements RERI; both assess interaction on the additive scale. Both measures suggest some degree of positive interaction, though neither reaches statistical significance in our data.

# Part II. Multiplicative Interaction (Product Term Approach)

2.1. Multiplicative Interaction

* **Definition:** Multiplicative interaction examines whether the joint effect of two exposures differs from what would be expected if their effects multiplied.
* **Formula:**
* **Interpretation:**
* Value = 1: No multiplicative interaction
* Value > 1: Super-multiplicative interaction
* Value < 1: Sub-multiplicative interaction

**1.** We calculate the multiplicative interaction using the previous output of logistic regression.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Exposure Combination** | **BMI** | **Mediterranean Diet** | **OR (vs. reference)** | **Symbol** |
| Neither exposure | 0 | 0 (High adherence) | 1.00 (reference) | OR₀₀ |
| High BMI only | 1 | 0 (High adherence) | 2.61 | OR₁₀ |
| Low MedDiet only | 0 | 1 (Low adherence) | 1.55 | OR₀₁ |
| Both exposures | 1 | 1 (Low adherence) | 3.79 | OR₁₁ |

The multiplicative interaction on the OR scale is:

**2.** In Stata, we use the **nlcom** command for this calculation:

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

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

hyper | Coefficient Std. err. z P>|z| [95% conf. interval]

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

m1\_ln\_mult\_int | -.0643341 1.014415 -0.06 0.949 -2.052551 1.923882

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

**3.**

. 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])

**4.**

. 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

**5.** The multiplicative interaction estimate (OR\_int = 0.94) is less than 1, suggesting a sub-multiplicative interaction between obesity and low Mediterranean diet adherence on hypertension risk. This means that the joint effect of these two risk factors appears to be less than what we would expect if their effects were perfectly multiplicative.

However, the 95% confidence interval (0.13, 6.85) includes 1, indicating that we cannot reject the null hypothesis of no multiplicative interaction at the α = 0.05 level. The wide confidence interval suggests considerable uncertainty in our estimate of the interaction effect. Therefore, while our point estimate suggests sub-multiplicativity, we do not have sufficient statistical evidence to conclude that the joint effects of obesity and low Mediterranean diet adherence deviate significantly from multiplicativity.

On the **OR scale**, exp(product term) = 0.94 (<1) hints sub-multiplicativity, but the CI is wide and crosses 1 ⇒ **no significant interaction**.

**6. Comparison of Interaction Measures in the Study of BMI and Mediterranean Diet on Hypertension**

1. **Multiplicative Interaction (OR scale):**

Point estimate (95% CI): 0.94 (0.13, 6.85) → sub-multiplicative

Interpretation: The joint effect appears to be slightly less than multiplicative

Statistical significance: Not significant (CI includes 1)

1. **Additive interaction (RERI on OR scale):**

Point estimate (95% CI): 0.63 (-2.48, 3.74) → super-additive

Interpretation: The joint effect appears to be greater than additive

Statistical significance: Not significant (CI includes 0)

1. **Key Insights from Both Measures:**

**Directionality Contrast:** The measures suggest different directions of interaction (sub-multiplicative but super-additive). This is not uncommon and highlights the value of reporting both scales.

**Statistical Uncertainty:**

* Both measures have wide confidence intervals
* Neither interaction is statistically significant
* The study may be underpowered to detect interactions

**Public Health Angle:** The positive RERI (0.63) may be more policy-relevant because it quantifies excess risk on the additive scale.

***Note:*** RERI/S computed from logistic models operate on the **OR** scale and approximate **risk-based** quantities only under the **rare-disease** assumption.

# Part III. Alternative Approaches

***3.1. Classic Interaction Analysis***

**1.**

. gen bmi\_nomd = bmi \* nomd

. label variable bmi\_nomd "Interaction: BMI × Low Mediterranean Diet Adherence"

**2.**

. 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

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

Note: \_cons estimates baseline odds.

RERI can be calculated as before using nlcom command in Stata:

. nlcom 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:

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

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

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 those previously reported, as expected.

# Part IV. Incorrect Coding and Reference Approaches

In epidemiological studies, variables must be coded so that higher values indicate increased risk [Knol et al., 2012]. When generating a four-level variable, the reference category must correspond to the absence of both risk factors (i.e., the lowest risk category).

Below we demonstrate the consequences of incorrect coding approaches:

***4.1 Wrong Reference (Incorrect Approach #1)***

**1.** **Using level 2 (BMI > 30 kg/m² with Mediterranean diet adherence) as baseline**

. char cat\_bmi\_nomd [omit] 2

**2.**

. xi: 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]

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

\_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

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

Note: \_cons estimates baseline odds.

**3.**

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

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

hyper | Coefficient 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

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

hyper | Coefficient 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

**4.** Problems with this approach:

**a.** Reference Category Misspecification:

* Using BMI > 30 kg/m² with Mediterranean diet adherence as baseline contradicts epidemiological principles
* The reference should represent the lowest risk category (absence of both risk factors)
* This misspecification leads to incorrect estimation of relative effects

**b.** Impact on Interaction Measures:

* RERI = 1.47 (95% CI: 0.23, 2.72) is misleading as it no longer measures excess risk above the minimum risk group
* Multiplicative interaction of 6.39 (95% CI: 0.88, 46.67) is incorrectly inflated
* Both measures lose their intended epidemiological interpretation

**Incorrect baseline (reference category)**

For additive interaction measures such as the RERI and the Synergy Index, the baseline category must correspond to the joint absence of both exposures (00). Choosing any other baseline makes the resulting OR11, OR10, and OR01 incomparable, because the contrasts no longer represent single-factor effects relative to “no exposure”.

To illustrate, if we (incorrectly) set the baseline to category 2 instead of 0 in i.cat\_bmi\_nomd, the estimated ORs change substantially and the resulting RERI becomes difficult or impossible to interpret in epidemiological terms. The numeric value can still be computed, but it no longer corresponds to the standard definition:

because the ORs are no longer referenced to (0,0).

Correct additive interaction analysis therefore requires coding the joint “neither exposure present” category as the base level.

***4.2. Reversed Coding (Incorrect Approach #2)***

The same problems arise if we follow a classic approach with the reference level of nomd indicating no dedication to the Mediterranean diet.

**1.**

. 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

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

**2.**

**a. Conceptual Issues:**

* When nomd is reverse-coded (err\_nomd), higher values indicate protection, not risk.
* The model becomes mixed: bmi represents **risk**, while err\_nomd represents **protection**.
* The product term (err\_bmi\_nomd) loses a clear epidemiologic interpretation because it combines opposing directions.

**b. Statistical Interpretation Challenges:**

* OR(err\_nomd) = **0.65** (<1) suggests protection; OR(bmi) = **2.45** (>1) indicates increased risk.
* The interaction OR for err\_bmi\_nomd = **1.07** is difficult to interpret because it aggregates effects with opposite coding.
* Direction and magnitude of the estimated interaction depend on coding choices; reversing coding can invert the apparent direction of effect.

**c. Practical Implications:**

* Public health messaging becomes complicated.
* Risk assessment and prevention strategies are more difficult to formulate.
* Results are not directly comparable to studies using standard risk-factor coding.

**Why it matters:** Consistent risk coding (**1 = higher risk**) and a reference group with **no exposures (00)** are essential for interpretable interaction analyses (both additive and multiplicative).

Without consistent coding, RERI and S lose their epidemiologic meaning, and comparisons across studies become unreliable.

**Incorrect reverse coding**

Additive interaction measures require that both exposures are coded in the same risk direction. If one variable is reversed (e.g., 1 = low risk, 0 = high risk) while the other is coded normally, the four joint exposure categories no longer align with the conceptual 00/10/01/11 structure used to define RERI and S.

In such cases, the fitted coefficients from the logistic model remain mathematically valid, but the quantities

no longer correspond to the intended causal ordering, making the resulting RERI and Synergy Index uninterpretable.

This is why both exposures should be coded so that the value “1” indicates higher hypothesised risk (or, at minimum, the *same* risk direction for both exposures). Inconsistent coding changes the meaning of the four exposure categories and invalidates epidemiological interpretation of additive interaction.

**Key lessons**

1. Use consistent **risk coding** for all exposures (1 = higher risk).
2. Set the **reference category to 00** (absence of both risk factors).
3. Report interaction on both scales (additive: RERI/S; multiplicative: product term), stating when OR-based measures are approximations (rare-outcome assumption).
4. Avoid mixing coefficients and ORs in reporting; use one scale and include CIs.
5. Re-code and re-fit when needed:
   * gen nomd = 1 - err\_nomd (restore risk direction)
   * gen bmi\_nomd = bmi \* nomd
   * logistic hyper bmi nomd bmi\_nomd and compute RERI/S via nlcom.