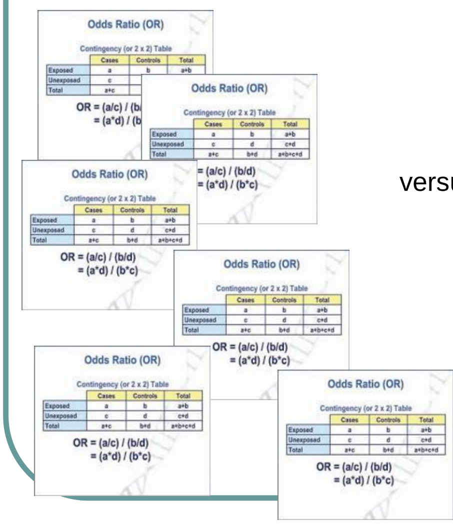
Interpretation of the logistic regression model

HL Chapter 3 – part 2

Confounding and interactions in logistic regression

Confounding

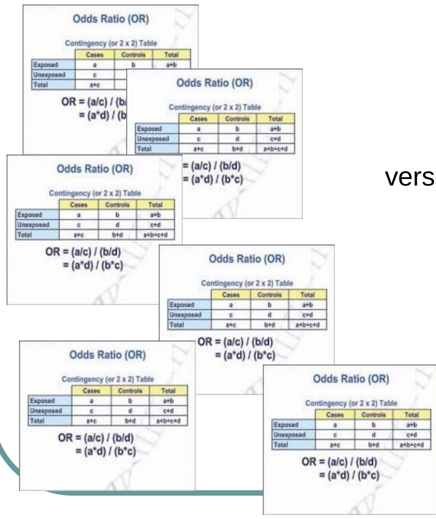
Confounding



versus

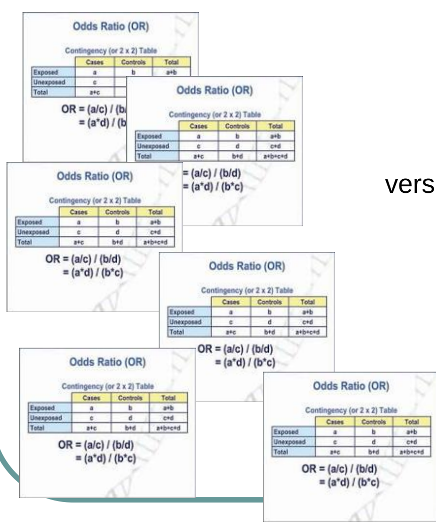
Interaction

Interaction - Multiplicative



versus

Interaction - Additive



versus

Goals

- Learn to investigate confounding using logistic regression
- Learn to investigate multiplicative interactions using logistic regression
- Learn to investigate additive interactions using a trick

Confounding - 2×2 table approach

1. Calculate crude 2×2 table and OR
2. Calculate stratum-specific 2×2 tables and ORs
3. Calculate adjusted OR = weighted (Mantel-Haenszel) average of stratum-specific ORs
4. Compare crude and adjusted OR using the 10% rule

Confounding - 2×2 table approach

- $\left|\frac{Crude\ OR - Adjusted\ OR}{Adjusted\ OR}\right| < about\ 10\%$ ➔ No evidence of confounding

- $\left|\frac{Crude\ OR - Adjusted\ OR}{Adjusted\ OR}\right| > about\ 10\%$ ➔ Evidence of confounding

- Note: Continuous confounders must be categorized

Confounding in logistic regression

Example: MYOPIA data set

Risk factor = gender

Potential confounder or effect modifier = SPHEQ

- Run the model without the potential confounder

```
proc logistic descending data=myopia;
  model myopic=gender;
run;
```

Confounding in logistic regression

Parameter	Coefficient	Pr > ChiSq	OR	95% Wald Confidence Limits	
Intercept	-2.0829	<.0001			
GENDER	0.3665	0.1274	**1.443**	0.901	2.311

This table was created by combining results from two tables in the SAS output

Confounding in logistic regression

- Run the model <u>with</u> the potential confounder
- Note: Continuous confounders do <u>not</u> have to be categorized
- Note: SPHEQ ranges from -0.699 to 4.372; therefore, a unit change of 1 (SAS default) is too big; use 0.1

Confounding in logistic regression

```
proc logistic descending data=myopia;
  model myopic=gender spheq/clodds=wald;
  units gender=1 spheq=0.1;
run;
```

Odds Ratio Estimates and Wald Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
GENDER	1.0000	**1.747**	0.999	3.055
SPHEQ	0.1000	0.681	0.627	0.739

Confounding in logistic regression

- $\left|\frac{Crude\ OR - Adjusted\ OR}{Adjusted\ OR}\right| = \left|\frac{1.443 - 1.747}{1.747}\right| = 17\%$
- There is evidence that SPHEQ is a confounder of GENDER

Continuous confounders

- Note: SPHEQ is the spherical equivalent refraction, a continuous variable
- What does adjusting for a continuous variable mean?
- Adjusting for a continuous variable means holding its value constant

Continuous confounders

- It doesn’t matter at which value of SPHEQ we compare males to females as long as we use the same value of SPHEQ for males and females
- This only works if there is no multiplicative interaction between GENDER and SPHEQ, i.e. if the OR for GENDER does not depend on the level of SPHEQ

10% rule vs. stat. significance

- Why do we use the 10% rule rather than statistical significance to assess confounding?
- Confounders must be associated with the outcome
  - Statistical significance addresses this
- Confounders must be associated with the risk factor of interest
  - Statistical significance does not address this

10% rule vs. stat. significance

- It can be shown that the "effect" of a risk factor is $\beta_1 + \beta_2(\bar{c_1} - \bar{c_0})$
- I.e. it is a combination of
  - The actual "effect" of the risk factor on the outcome (β_1) AND
  - The "effect" due to confounding ($\beta_2(\bar{c_1} - \bar{c_0})$)
- $\bar{c_0}$= average of the confounder among those not exposed to the risk factor
- $\bar{c_1}$= average of the confounder among those exposed to the risk factor

10% rule vs. stat. significance

- The "effect" due to confounding, is the product of
  - The "effect" of the confounder on the outcome (β_2) and
  - The difference between the average of the confounder among the exposed and the unexposed ($\overline{c_1} - \overline{c_0}$)
- Confounding does not only depend on the significance of the confounder, i.e. the significance of β_2
- Even if β_2, the age coefficient, is non-significant, confounding may be present if $\overline{a_1} - \overline{a_0}$ is large

Multiplicative interaction - 2×2 table approach

- Stratum-specific 2×2 tables and ORs
  - ORs similar ➔ No multiplicative interaction
  - OR in exposed stratum greater
    ➔ Synergistic multiplicative interaction
  - OR in unexposed stratum greater
    ➔ Antagonistic multiplicative interaction

- Note: Continuous effect modifiers must be categorized

Multiplicative interaction – 4-row table

- $OR_{both} < OR_{one} \times OR_{other}$ ➔ Antagonistic multiplicative interaction
- $OR_{both} = OR_{one} \times OR_{other}$ ➔ No multiplicative interaction
- $OR_{both} > OR_{one} \times OR_{other}$ ➔ Synergistic multiplicative interaction

- Note: Continuous effect modifiers must be categorized

Multiplicative interaction in logistic regression

- The logistic regression model is a multiplicative model and checks for multiplicative interactions

- If additive interactions are of interest, we can use a linear link model (HL Chapter 10.9)

Multiplicative interaction in logistic regression

Example: MYOPIA example continued

- Recall: Adjusting GENDER for SPHEQ means comparing males to females at the same value of SPHEQ
- This only works if there is no multiplicative interaction between GENDER and SPHEQ, i.e. if the OR for GENDER does not depend on the level of SPHEQ

Multiplicative interaction in logistic regression

```
proc logistic descending data=myopia;
  model myopic=gender spheq gender*spheq;
run;
```

Parameter	Coefficient	Pr > ChiSq
Intercept	-0.1911	0.5240
GENDER	0.4916	0.2369
SPHEQ	-3.9483	<.0001
GENDER*SPHEQ	0.1850	**0.8261**

Multiplicative interaction in logistic regression

- There is no significant multiplicative interaction between GENDER and SPHEQ ($p=0.8261$)

- Note: Testing the significance of interaction terms often suffers from low power. Therefore, an α level of 0.1 is often used instead of the traditional 0.05

Interpretation of the OR in the presence of confounding

- Interaction not statistically significant
  ➔ drop interaction term

- Back to this table…

Odds Ratio Estimates and Wald Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
GENDER	1.0000	**1.747**	0.999	3.055
SPHEQ	0.1000	**0.681**	0.627	0.739

Interpretation of the OR in the presence of confounding

- After adjusting for SPHEQ (i.e. holding SPHEQ constant), females are about 1.75 times as likely to be myopic as males (75% more likely to be myopic than males)
- After adjusting for GENDER (i.e. holding GENDER constant), an increase in SPHEQ of 0.1 results in a 32% decreased risk of myopia

Interpretation of the OR in the presence of multiplicative interaction

Example: GLOW500 data set

Risk factor = PRIORFRAC

Potential confounder or effect modifier = Age

```
proc logistic descending data=glow500;
  model fracture=priorfrac; run;
```

Parameter	Coefficient	Pr > ChiSq	OR	95% Wald Confidence Limits	
Intercept	-1.4167	<.0001			
PRIORFRAC	1.0638	<.0001	**2.897**	1.871	4.486

Interpretation of the OR in the presence of multiplicative interaction

```
proc logistic descending data=glow500;
  model fracture=priorfrac age/clodds=wald;
  units priorfrac=1 age=10 ;
run;
```

Odds Ratio Estimates and Wald Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
PRIORFRAC	1.0000	**2.314**	1.462	3.661
AGE	10.0000	1.510	1.189	1.917

Interpretation of the OR in the presence of multiplicative interaction

- $\left|\frac{Crude\ OR - Adjusted\ OR}{Adjusted\ OR}\right| = \left|\frac{2.897 - 2.314}{2.314}\right| = 25\%$
- There is evidence that AGE is a confounder of PRIORFRAC
- Note: Adjusting for age means holding age constant

Interpretation of the OR in the presence of multiplicative interaction

- For this interpretation to be valid, there cannot be evidence of multiplicative interaction

```
proc logistic descending data=glow500;
  model fracture=priorfrac age priorfrac*age;
run;
```

Parameter	Estimate	Pr > ChiSq
Intercept	-5.6893	<.0001
PRIORFRAC	4.9612	0.0061
AGE	0.0625	<.0001
PRIORFRAC*AGE	-0.0574	0.0218

Evidence of multiplicative interaction

Interpretation of the OR in the presence of multiplicative interaction

- There is evidence of multiplicative interaction between PRIORFRAC and AGE
- The OR of PRIORFRAC depends on age
- Presenting just one "average" adjusted OR doesn't make sense
- How do we calculate ORs in the presence of interactions?
  - Logit differences and contrast statements

IMPORTANT

In the presence of multiplicative interaction, it does not make sense to present just one OR

OR for PRIOFRAC yes (1) vs. no (0) at age *a*

- Logit = $g(PRIORFRAC, AGE\ )$

$$= \beta_0 + \beta_1 PRIORFRAC + \beta_2 AGE + \beta_3 PRIORFRAC \times AGE$$

- Logit difference =

$$g(PRIORFRAC = 1, AGE = a\ )$$
$$- g(PRIORFRAC = 0, AGE = a\ )$$

$$= (\beta_0 + \beta_1 1 + \beta_2 a + \beta_3 1 \times a) - (\beta_0 + \beta_1 0 + \beta_2 a + \beta_3 0 \times a)$$
$$= \beta_1 + a\beta_3 \rightarrow OR = e^{\beta_1 + a\beta_3}$$

Plug in different ages for a, e.g., 55, 65, 75, 85

OR for PRIOFRAC yes (1) vs. no (0) at age *a*

- $OR = e^{\beta_1 + a\beta_3}$

Coefficient	Corresponding model covariate
β_1	PRIORFRAC
β_2	AGE
β_3	PRIORFRAC *AGE

Age	OR	SAS
55	$e^{\beta_1+55\beta_3}$	priorfrac 1 age 0 priorfrac*age 55
65	$e^{\beta_1+65\beta_3}$	priorfrac 1 age 0 priorfrac*age 65
75	$e^{\beta_1+75\beta_3}$	priorfrac 1 age 0 priorfrac*age 75
85	$e^{\beta_1+85\beta_3}$	priorfrac 1 age 0 priorfrac*age 85

OR for a 10 year increase in age at PRIORFRAC=*p*

- Logit = $g(PRIORFRAC, AGE\ )$

$$= \beta_0 + \beta_1 PRIORFRAC + \beta_2 AGE + \beta_3 PRIORFRAC \times AGE$$

- Logit difference =

$$g(PRIORFRAC = p, AGE = a + 10\ ) - g(PRIORFRAC = p, AGE = a\ )$$

$$= (\beta_0 + \beta_1 p + \beta_2(a+10) + \beta_3 p \times (a+10)) - (\beta_0 + \beta_1 p + \beta_2 a + \beta_3 p \times a)$$

$$= 10\beta_2 + 10p\beta_3 \Rightarrow OR = e^{10\beta_2 + 10p\beta_3}$$

$\boxed{\text{Plug in different values for } p\text{, i.e. 0, 1}}$

OR for a 10 year increase in age at PRIORFRAC=*p*

- $OR = e^{10\beta_2 + 10p\beta_3}$

Coefficient	Corresponding model covariate
β_1	PRIORFRAC
β_2	AGE
β_3	PRIORFRAC *AGE

PRIOR FRAC	OR	SAS contrasts
0	$e^{10\beta_2 + 10p\beta_3}$	priorfrac 0 age 10 priorfrac*age 0
1	$e^{10\beta_2 + 10p\beta_3}$	priorfrac 0 age 10 priorfrac*age 10

SAS code

```
proc logistic descending data=glow500;
  model fracture=priorfrac age priorfrac*age;

contrast 'Priorfrac 1 vs 0, age=55'
                   priorfrac 1 age 0 priorfrac*age 55 /estimate=exp;
contrast 'Priorfrac 1 vs 0, age=65'
                   priorfrac 1 age 0 priorfrac*age 65/estimate=exp;
contrast 'Priorfrac 1 vs 0, age=75'
                   priorfrac 1 age 0 priorfrac*age 75/estimate=exp;
contrast 'Priorfrac 1 vs 0, age=85'
                   priorfrac 1 age 0 priorfrac*age 85/estimate=exp;
```

Continued on next slide

SAS code

Continued from previous slide

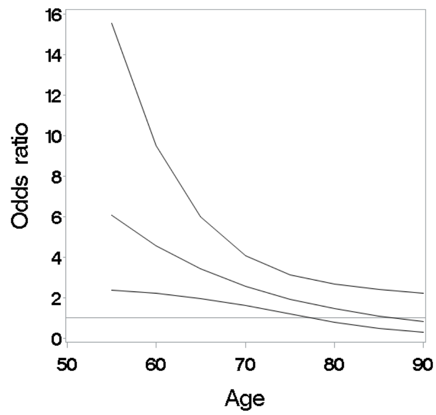
```
contrast '10 year age increase, priorfrac=0'
        priorfrac 0 age 10 priorfrac*age 0 /estimate=exp;
contrast '10 year age increase, priorfrac=1'
        priorfrac 0 age 10 priorfrac*age 10 /estimate=exp;
run;
```

OR for PRIOFRAC yes (1) vs. no (0) at age *a*

Contrast	OR	Confidence Limits		Pr > ChiSq
Priorfrac 1 vs 0, age=55	6.0818	2.3816	15.5307	0.0002
Priorfrac 1 vs 0, age=65	3.4263	1.9593	5.9918	<.0001
Priorfrac 1 vs 0, age=75	1.9303	1.1993	3.1069	0.0068
Priorfrac 1 vs 0, age=85	1.0875	0.4944	2.3921	0.8348

OR for PRIOFRAC yes (1) vs. no (0) at age *a*

- Graph of ORs and 95% CIs for PRIORFRAC=1 vs. 0 by age



Interpretation of the results

- At age 55, persons with a prior facture are about 6 times as likely to have a fracture in the first year than persons without a prior fracture. This increased risk is statistically significant.
- As age increases, the effect of prior fracture on fracture in first year decreases
- At about age 78, the increased risk is no longer statistically significant
- Above age 80 no increased risk is observed

OR for a 10 year increase in age at PRIORFRAC=*p*

Contrast	OR	Confidence Limits		Pr > ChiSq
10 year age increase, priorfrac=0	1.8685	1.3800	2.5298	<.0001
10 year age increase, priorfrac=1	1.0527	0.7160	1.5476	0.7941

Interpretation of the results

- Among persons without a prior fracture, a 10 year increase in age increases the risk of fracture in first year by about 86%; the increase is statistically significant
- Among persons with a prior fracture, a 10 year increase in age does not significantly increases the risk of fracture in first year

Additive interactions in logistic regression

- Example: Myopia data set
- Logistic regression of MYOPIC on
  - DADMY
  - SPHEQ_50 (1 if SPHEQ≤0.5 and 0 if SPHEQ>0.5)

Additive interactions in logistic regression

Parameter	Estimate	Pr > ChiSq
Intercept	-3.4701	<.0001
DADMY	0.8310	0.0790
spheq_50	2.2661	<.0001
DADMY*spheq_50	0.0852	**0.8834**

- Interaction p-value much greater than 0.1

  ➔ no evidence of multiplicative interaction

4-row table from logistic regression

- If the interaction variables are dichotomous, a 4-row table can be created
- For the logistic regression model

$$g(DADMY, SPHEQ_{50})$$

$$= \beta_0 + \beta_1 DADMY + \beta_2 SPHEQ_{50}$$
$$+\beta_3 DADMY \times SPHEQ_{50}$$

4-row table from logistic regression

DADMY	spheq_50	$\widehat{OR} = e^{\hat{g}}$, where $\hat{g} =$	SAS contrasts
1	1	$\hat{g}(DADMY = 1, SPHEQ_{50} = 1)$ $=\hat{\beta}_0 + \hat{\beta}_1(1) + \hat{\beta}_2(1) + \hat{\beta}_3(1 \times 1)$	dadmy **1** spheq_50 **1** dadmy*spheq_50 **1**
1	0	$\hat{g}(DADMY = 1, SPHEQ_{50} = 0)$ $=\hat{\beta}_0 + \hat{\beta}_1(1) + \hat{\beta}_2(0) + \hat{\beta}_3(1 \times 0)$	dadmy **1** spheq_50 **0** dadmy*spheq_50 **0**
0	1	$\hat{g}(DADMY = 0, SPHEQ_{50} = 1)$ $=\hat{\beta}_0 + \hat{\beta}_1(0) + \hat{\beta}_2(1) + \hat{\beta}_3(0 \times 1)$	dadmy **0** spheq_50 **1** dadmy*spheq_50 **0**
0	0	*Reference category*	

4-row table from logistic regression

```
proc logistic descending data=myopia;
  model myopic=dadmy spheq_50 dadmy*spheq_50;

  contrast 'both vs. neither'  dadmy 1 spheq_50 1
                               dadmy*spheq_50 1/estimate=exp;
  contrast 'DADMY vs. neither' dadmy 1 spheq_50 0
                               dadmy*spheq_50 0/estimate=exp;
  contrast 'SPHEQ vs. neither' dadmy 0 spheq_50 1
                               dadmy*spheq_50 0/estimate=exp;
run;
```

4-row table from logistic regression

- Multiplicative interaction? $9.64 \times 2.3 = 22.2 \approx 24.1$
  - ➔ No evidence of multiplicative interaction
- Additive interaction? $9.64 + 2.3 - 1 = 10.94 < 24.1$
  - ➔ Evidence of synergistic additive interaction

DADMY	spheq_50	Contrast	**OR**	Confidence Limits	
1	1	both vs. neither	**24.1049**	10.2825	56.5087
1	0	DADMY vs. neither	**2.2957**	0.9083	5.8025
0	1	SPHEQ vs. neither	**9.6420**	3.8491	24.1530
0	0		**1.00**		

4-row table from logistic regression

- **Remember that all ORs must be greater than 1 for this method to work**

Additive interactions in an appropriate regression model

- To create an additive model for these data, use the linear link binomial model
- Advantage: Allows testing for additive interactions
- Disadvantage:
  - The linear link model can result in values of $\hat{\pi}$ that are greater than 1 or less than 0
  - In this case the model cannot be used to test for additive interactions

Fitting the linear link model

```
proc genmod descending data=myopia;
  model myopic = dadmy spheq_50 dadmy*spheq_50
                 / dist=bin link = identity;
  output out=pdat p=phat;
run;
```

Linear link

Binomial distribution (0/1 outcome)

Save $\hat{\pi}$ to make sure $0 <= \hat{\pi} <= 1$

```
proc univariate data=pdat; var phat; run;
```

Check min and max of $\hat{\pi}$

Results from proc genmod

Parameter	Coefficient	Pr > ChiSq
Intercept	0.0302	0.0072
DADMY	0.0365	0.0758
spheq_50	0.2006	<.0001
DADMY*spheq_50	0.1613	**0.0253**

- Interaction p-value much less than 0.1
  ➔ evidence of additive interaction

Results from proc univariate

- Minimum of $\hat{\pi}$ is >0
- Maximum of $\hat{\pi}$ is <1

Quantile	Estimate
100% Max	0.4285714
0% Min	0.0301724

Complete analysis MYOPIA data set

Step 1: Data step

```
libname sdat 'C:\ERHS642';

data myopia;
  set sdat.myopia;
          if -1<=spheq<=0.5 then spheq_50=1;
   else if        spheq>  0.5 then spheq_50=0;
run;
```

Step 2: Confounding

```
proc logistic descending data=myopia;
   model myopic=dadmy;
run;

proc logistic descending data=myopia;
  model myopic=dadmy spheq_50;
run;
```

Step 2: Confounding, cont.

Effect	OR	95% Wald Confidence Limits	
DADMY	2.533	1.535	4.181

Effect	OR	95% Wald Confidence Limits	
DADMY	2.430	1.417	4.166
spheq_50	10.195	5.921	17.554

- |(2.533-2.430)/2.430 |≈4%

  ➔ no evidence of confounding

Step 3: Multiplicative interaction

```
proc logistic descending data=myopia;
  model myopic = dadmy spheq_50
                                  dadmy*spheq_50;
run;
```

Step 3: Multiplicative interaction, cont.

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.4701	0.3838	81.7549	<.0001
DADMY	1	0.8310	0.4731	3.0857	0.0790
spheq_50	1	2.2661	0.4685	23.3944	<.0001
DADMY*spheq_50	1	0.0852	0.5811	0.0215	**0.8834**

- Interaction p-value much greater than 0.1
  - ➔ no evidence of multiplicative interaction

Step 4: Interactions based on 4-row table

```
proc logistic descending data=myopia;
  model myopic=dadmy spheq_50 dadmy*spheq_50;

  contrast 'both vs. neither'  dadmy 1 spheq_50 1
                               dadmy*spheq_50 1/estimate=exp;
  contrast 'DADMY vs. neither' dadmy 1 spheq_50 0
                               dadmy*spheq_50 0/estimate=exp;
  contrast 'SPHEQ vs. neither' dadmy 0 spheq_50 1
                               dadmy*spheq_50 0/estimate=exp;
run;
```

Step 4: Interactions based on 4-row table, cont.

DADMY	spheq_50	Contrast	**OR**	Confidence Limits	
1	1	both vs. neither	**24.1049**	10.2825	56.5087
1	0	DADMY vs. neither	**2.2957**	0.9083	5.8025
0	1	SPHEQ vs. neither	**9.6420**	3.8491	24.1530
0	0		**1.00**		

- Multiplicative interaction? $9.64 \times 2.3 = 22.2 \approx 24.1$
  ➔ No evidence of multiplicative interaction
- Additive interaction? $9.64 + 2.3 - 1 = 10.94 < 24.1$
  ➔ Evidence of synergistic additive interaction

Step 5: Additive interaction based on the linear link model

```
proc genmod descending data=myopia;
  model myopic = dadmy spheq_50 dadmy*spheq_50/
                                          dist=bin link = identity;
  output out=pdat p=phat;
run;

proc univariate data=pdat;
  var phat;
run;
```

Step 5: Results from proc genmod

Parameter	Coefficient	Pr > ChiSq
Intercept	0.0302	0.0072
DADMY	0.0365	0.0758
spheq_50	0.2006	<.0001
DADMY*spheq_50	0.1613	**0.0253**

- Interaction p-value much less than 0.1
  ➔ evidence of additive interaction

Step 5: Results from proc univariate

- Minimum of $\hat{\pi}$ is >0
- Maximum of $\hat{\pi}$ is <1

Quantile	Estimate
100% Max	0.4285714
0% Min	0.0301724

Step 6: Interpretation

- Confounding
  - There is no evidence that SPHEQ_50 is a confounder of DADMY
- Multiplicative interaction
  - There is no evidence of multiplicative interaction between SPHEQ_50 and DADMY
  - The OR of DADMY is not significantly different for subjects with SPHEQ_50≤0.5 and for subjects with SPHEQ_50>0.5

Step 5: Interpretation, cont.

- Additive interaction
  - There is evidence of synergistic additive interaction
  - The risk difference of DADMY is significantly different for subjects with SPHEQ_50≤0.5 and SPHEQ_50>0.5
  - It appears that DADMY and SPHEQ_50 do not act independently