

Course on :

Data Analysis in

# CASE CONTROL Studies

دوره آموزشی:

تحلیل آماری در

مطالعات مورد شاهدهی



Exposed



Not exposed

آغاز دوره از  
سه شنبه ۲۹ مهر

۱۸ ساعت  
آموزش  
نظری و  
عملی

طراحی و تدریس: دکتر عباس کشتکار

عضو هیات علمی دانشگاه علوم پزشکی تهران

مدرس بیش از ۱۵۰ عنوان دوره و کارگاه آموزشی برای پژوهشگران

زمان جلسات

سه شنبه شبها: ۸ تا ۱۱

# گامهای تحلیل آماری در مطالعات مورد شاهدهی :

1. تحلیلهای توصیفی یا Descriptive Analysis

2. تحلیل تک متغیره یا Univariable A.

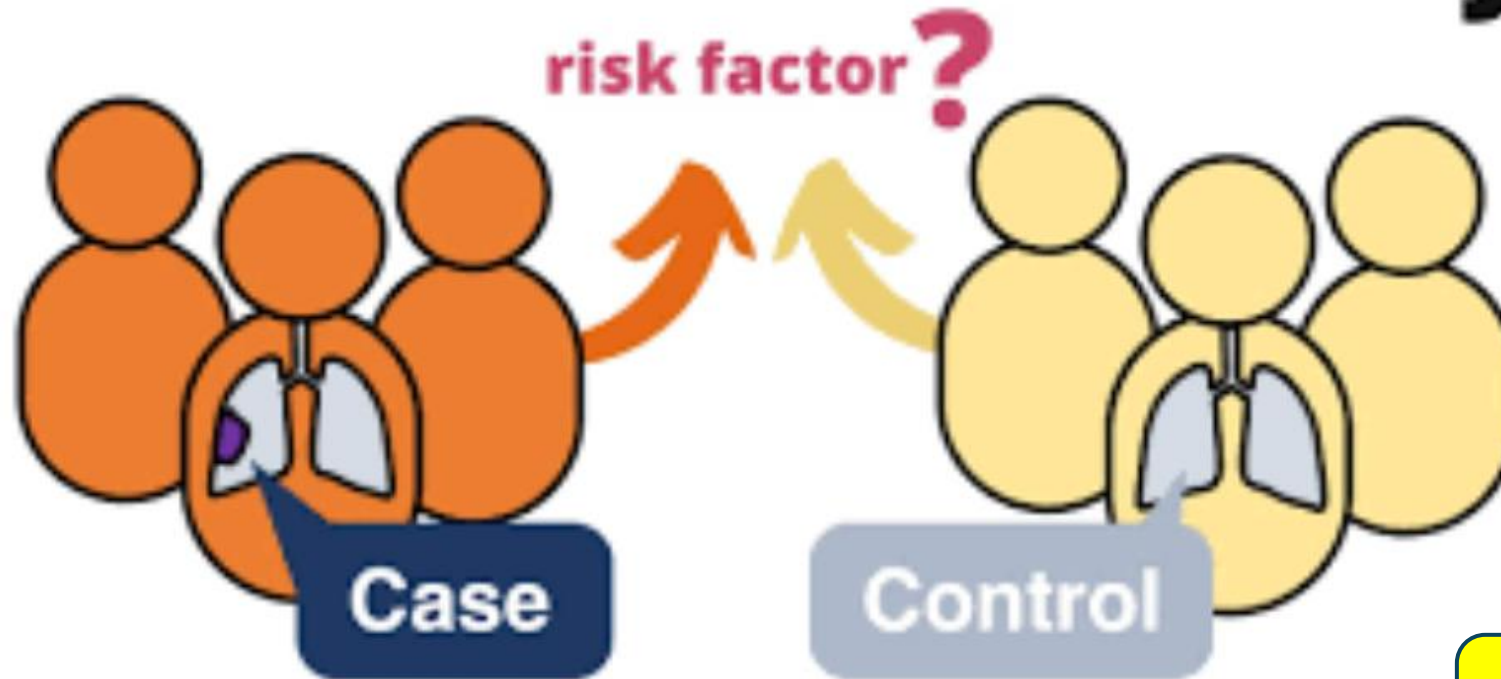
3. تحلیل چند متغیره یا Multivariable A.

4. تحلیلهای اضافی یا Additional A شامل:

(a) تحلیلهای زیر گروه یا Sub-group A.

(b) تحلیلهای حساسیت یا Sensitivity A.

# Case-control study



درس چهارم

تحليل چندمتغیره (تصحیح شده) در  
مطالعات مورد شاهدهی

# متدهای تحلیل چندمتغیره در مطالعات مورد شاهدهی :

متدهای زیر در مطالعات مورد شاهدهی، بعنوان «روشهای چندمتغیره»، کاربرد دارند :

• روش طبقه‌بندی یا مانتل هنسل (Mantel-Haenszel)

• روش رگرسیون لجستیک (معمولی یا شرطی)

# تعریف عملیاتی روش «رگرسیون لجستیک»: (۱)

روش «رگرسیون لجستیک» یک روش با ماهیت «چندمتغیره» (و البته تک متغیره) است که مانند بسیاری از روشهای رگرسیونی، بر اساس ویژگیهای زیر طبقه‌بندی میشود:

• ماهیت پیامد (متغیر وابسته):

• یک متغیر دوسطحی یا Binary: در مطالعات مورد شاهدهی، گروه مورد و شاهد، بترتیب کدهای یک و صفر، میگیرند.

• ماهیت متغیر / متغیرهای مستقل (Predictors):

- میتواند هر نوع متغیری باشد (کمی، باینری، چندسطحی رتبه‌ای، چندسطحی اسمی)
- از نظر تعداد متغیرها نیز محدودیتی وجود ندارد (یک یا چند متغیر ...)

# تعریف عملیاتی روش «رگرسیون لجستیک»: (۲)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$$

مدل عمومی رگرسیون

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p)}}$$

مدل رگرسیون لجستیک  
(پیامد احتمال)

$$\text{logit}(p) = \ln \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

مدل رگرسیون لجستیک  
(پیامد لجیت)

# تعریف عملیاتی روش «رگرسیون لجستیک»: (۳)

مهمترین نکات در کاربرد و تفسیر «رگرسیون لجستیک»، عبارتند از :

## • نکات کاربردی مهم (۱):

- متغیر وابسته (پیامد)، باید بصورت یک متغیر Binary (دو وضعیتی) با کدهای «صفر» (بعنوان گروه کنترل) و «یک» (بعنوان گروه کیس یا مورد)، تعریف گردد.
- جایگاه متغیر «مواجهه اصلی» و «مخدوشگر / کوواریت» در مدل رگرسیون، تفاوتی با یکدیگر نداشته و هر یک از آنها در جایگاه X ها قرار میگیرند. (مانند اثر آزیست در وقوع سرطان ریه، بعد از تصحیح اثر سن بعنوان مخدوشگر)
- جایگاه «برهمکنش» (اینتراکشن) بین دو متغیر (مانند مصرف سیگار و مواجهه با آزیست در وقوع کانسر ریه)، باید بصورت Multiplicative Interaction (برهمکنش ضربی)، تعریف گردد.

# تعریف عملیاتی روش «رگرسیون لجستیک»: (۴)

چگونه متغیر متغیرهای مستقل و مخدوشگرها / کوواریتها را انتخاب نماییم

## • نکات کاربردی مهم:

- مبتنی بر مرور متون، برخلاف تصور بسیاری از محققین، انتخاب متغیرهای فوق بصورت آماری (معنی داری آماری) نبوده و باید براساس Priori Knowledge انتخاب گردد.
- اهمیت پدیده Collinearity (برهمخطی) همبستگی بسیار قوی بین حداقل یک زوج از متغیرهای مستقل / مخدوشگر (پرديكتورها) که موجب تخریب مدل رگرسیون و بخصوص افزایش بسیار بالای واریانس برآوردها خواهد شد.
- پرهیز از انتخاب برخی متغیرها متغیرهایی که بخصوص علامت یا نشانه یک بیماری باشند مانند سرفه در موارد بیماری کانسر ریه، درد دندان در موارد پوسیدگی دندان و نظایر آنها ...

# تعریف عملیاتی روش «رگرسیون لجستیک»: (۵)

مهمترین نکات در کاربرد و تفسیر «رگرسیون لجستیک»، عبارتند از:

• نکات کاربردی مهم (۲): ماهیت شاخص «ضریب رگرسیون» یا «بتا»

دقت داشته باشید که «بتا» در دو مقیاس ارائه میگردد:

• مقیاس «لگاریتمی»: معمولاً همانند بقیه مدل‌های رگرسیونی به Coefficient معروف است و در

تفسیر مقدار عددی نتایج (بعنوان اندازه اثر)، کاربردی ندارد، اما معنی‌داری آماری و گزارش  $p$  value آن با مقیاس دوم، تفاوتی ندارد.

• مقیاس «اکسپوننشIAL» (وقعی / عملی): این مقیاس که با Exponentiate مقدار بتا در مقیاس

قبلی، حاصل میشود، همان شاخص Odds Ratio (OR) است و مطابق آن تفسیر میگردد.

# تعریف عملیاتی روش «رگرسیون لجستیک»: (۶)

مهمترین نکات در کاربرد و تفسیر «رگرسیون لجستیک»، عبارتند از :

## • نکات کاربردی مهم (۳) : تفسیر شاخص «OR» یا «Exp. Beta»

- برای یک X کمی (مانند سن): هر یک واحد افزایش در متغیر X، شانس وقوع پیامد را به میزان B برابر (یا درصد)، افزایش / کاهش میدهد.
- برای یک X باینری (مانند جنس): در گروه با کد بزرگتر در متغیر X، در مقایسه با گروه با کد کوچکتر، شانس وقوع پیامد، به میزان B برابر (یا درصد)، افزایش / کاهش می‌یابد.
- برای یک X چندسطحی اسمی / رتبه‌ای (مانند قومیت یا وضعیت اجتماعی اقتصادی): در هر یک از افراد با کدهای مختلف متغیر X، در مقایسه با کمترین کد عددی (گروه مرجع) آن متغیر، شانس وقوع پیامد، به میزان B برابر (یا درصد) افزایش / کاهش می‌یابد.

# تعریف عملیاتی روش «رگرسیون لجستیک»: (۷)

مثال عملی برای تفسیر خروجی «رگرسیون لجستیک»:

- **Example:** Outcome; MI

- $X_1$ : smoker (binary)
- $X_2$ : age (continuous)
- $X_3$ : HTN (binary)

| Variable | Coefficient            |
|----------|------------------------|
| Constant | $-4.50 = \hat{\alpha}$ |
| $X_1$    | $0.70 = \hat{\beta}_1$ |
| $X_2$    | $0.05 = \hat{\beta}_2$ |
| $X_3$    | $0.42 = \hat{\beta}_3$ |

- **Interpretation:**

- Odds of MI in the smoker subjects is ( $e^{0.70} \approx 2.01$ ) times of non-smoker subjects
- Odds of MI for 1 year aging of subjects increases ( $e^{0.05} \approx 1.05$ ) times (5% increasing Odds of MI)
- Odds of MI in the hypertensive subjects is ( $e^{0.42} \approx 1.52$ ) times of normotensive subjects (increases odds of MI by 52% )

# تعریف عملیاتی روش «رگرسیون لجستیک»: (۸)

## نکات چالشی در مدل‌های رگرسیون لجستیک در مطالعات مورد شاهدهی

- یک مدل نهایی یا **Final Model** چقدر میتواند با ویژگیهای یک مدل رگرسیون لجستیک، منطبق باشد؟  
(مفهوم برازش مدل)
- یک مدل نهایی یا **Final Model** به چه میزان میتواند پیامد مورد نظر را درست پیش‌بینی نماید؟ (مفهوم عملکرد مدل)
- دو سوال فوق عمدتاً در کاربرد مدل‌های رگرسیونی (هر مدل رگرسیون و به تبع آن، مدل‌های رگرسیون لجستیک) در «پیش‌بینی» یا **Prediction modeling Studies** مطرح هستند و چون مطالعات مورد شاهدهی، برای اهداف «مدل‌های پیش‌بینی» مناسب نیستند، بسیاری از متخصصین اعتقاد دارند که پاسخ به این دو سوال در مدل‌های رگرسیون لجستیک در این مطالعات، کاربردی ندارند، اما برخی داوران آماری از شما این دو سوال را میپرسند!

# تعریف عملیاتی روش «رگرسیون لجستیک»: (۹)

چگونه به دو سوال زیر در مدل‌های رگرسیون لجستیک، پاسخ دهیم؟

- یک مدل نهایی یا Final Model چقدر میتواند با ویژگی‌های یک مدل رگرسیون لجستیک، منطبق باشد؟ (مفهوم برازش مدل): با آماره Hosmer-Lemshaw و P value مربوطه

- یک مدل نهایی یا Final Model به چه میزان میتواند پیامد مورد نظر را درست پیش‌بینی نماید؟ (مفهوم عملکرد مدل): با شاخص Nagelkerke R-squared و یا سهم افرادی که در گروه‌های پیامد مثبت و منفی، درست پیش‌بینی شده‌اند یا Correct-classified Proportion

# نکات مهم در کاربرد رگرسیون لجستیک در Stata

- ✓ ساختار دیتا در بکاربردن «رگرسیون لجستیک» در دو دسته مطالعات «غیرهمسان» و «همسان»، هر دو بصورت Long format است (گروه مورد و شاهد، در رکوردها یا ردیفهای جداگانه)
- ✓ متغیر پیامد (مورد و شاهد) با کدهای ۱ و ۰، متغیرهای کمی با مقیاس واقعی و برای متغیرهای دو یا چند سطحی نیز از کدهای حداقل صفر یا ۱، میتوانید استفاده نمایید.
- ✓ دستورات استیتا برای روشهای رگرسیونی لجستیک، بدین شرح است :
  - رگرسیون لجستیک کلاسیک یا غیرشرطی (برای مطالعات غیرهمسان): با دستور **“logit”** یا **“logistic”**
  - رگرسیون لجستیک شرطی (برای مطالعات همسان فردی): با دستور **“clogit”**، که این دستور نیازمند متغیر همسان یا **Matching Variable** است.

# کار عملی ۱: تحلیل رگرسیون لجستیک در یک مطالعه مورد شاهدی غیرهمسان

در دیتاست LBW\_unmatched CCS.dta که یک مطالعه غیرهمسان با نسبت تقریبی شاهد به مورد ۲ به ۱ است، اهداف زیر را ارزیابی نمایید:

✓ ارزیابی «رابطه بین سن مادران (بصورت کمی) با وقوع تولد نوزاد کم وزن» و تفسیر خروجی

✓ ارزیابی «رابطه بین وزن مادران (یکبار بصورت کمی و یکبار نیز دووضعیتی) با وقوع تولد نوزاد کم وزن» و مقایسه و تفسیر خروجی

✓ ارزیابی نقش برهمکنشی بین «نژاد (بصورت دووضعیتی غیرسفید در مقابل سفید) و مصرف سیگار»، و تفسیر خروجی و تعیین بهترین راهکار گزارش این هدف

✓ ارزیابی نقش برهمکنشی بین «سن مادران (بصورت کمی) و مصرف سیگار» و تفسیر خروجی ...

## کار عملی ۲: تحلیل رگرسیون لجستیک در یک مطالعه مورد شاهدی همسان

در دیتاست LBW2\_Matched CCS.dta که یک مطالعه همسان روی متغیر سن مادران با نسبت شاهد به مورد ۱ به ۱ است، اهداف زیر را ارزیابی نمایید:

✓ ارزیابی «رابطه بین وزن مادران (بصورت کمی) با وقوع تولد نوزاد کم وزن» و تفسیر خروجی

✓ ارزیابی «رابطه بین مصرف سیگار با وقوع تولد نوزاد کم وزن» و تفسیر خروجی

✓ ارزیابی «رابطه بین مصرف سیگار بعد از تصحیح اثر مخدوشگری بالقوه نژاد (غیر سفید در مقابل سفید)»، و تفسیر خروجی

✓ ارزیابی نقش برهمکنشی بین «وزن مادران (بصورت کمی) و مصرف سیگار» و تفسیر خروجی ...

مشاهده لیست دوره‌های آموزشی [www.researchware.org/shop](http://www.researchware.org/shop)

**پایان بخش دوم درس چهارم**  
**این دوره ادامه دارد ... !**

برای کسب اطلاع از آخرین اخبار آموزشی کلیه دوره های آموزشی آکادمی پژوهش‌افزار (دکتر کشتکار)،  
در گروه تلگرامی زیر عضو شوید: @amarafzar2011