Multivariate Analysis의 확장: 상호작용 그리고 곡선형 모형

이 정 복

울산대학교 의과대학 아산병원 신경과

Jung Bok Lee, MD, PhD

Department of Clinical Epidemiology and Biostatistics, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

Fundamental Concepts

(Included 2013 Spring Talks Slides)

서울아산병원 Asan Medical Center 2

Risk Factor (as well prognostic, predictive)

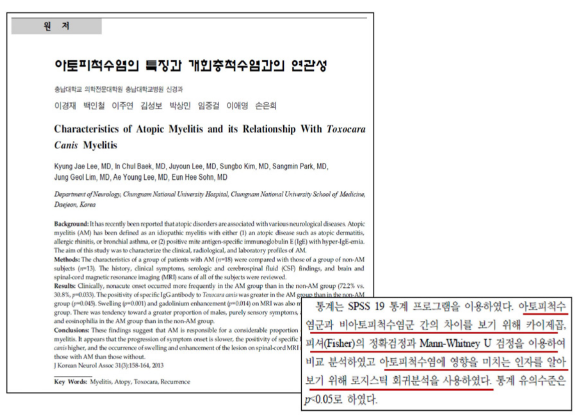
- Definition: A **risk factor** is a variable associated with an increased risk of disease or infection. Sometimes, **determinant** is also used, being a variable associated with either increased or decreased risk(Wikipedia).
- **Independent** Risk Factor:
  Outcome ← Candidate Risk Factors(A & B) : correlated or causal
  Risk Factor A ⊥ Risk Factor B
- Interpretation: If factor A is an independent risk factor of an outcome, factor A should be associated with the outcome regardless of low & high level of factor B.

서울아산병원 Asan Medical Center

Univariate & Multivariate Analysis

- **Univariate analysis** is the simplest form of quantitative (statistical) analysis. The analysis is carried out with the description of a single variable in terms of the applicable unit of analysis.
- Univariate analysis contrasts with bivariate analysis – the analysis of two variables simultaneously – or multivariate analysis – the analysis of multiple variables simultaneously. Univariate analysis is commonly used in the first, descriptive stages of research, before being supplemented by more advanced, inferential bivariate or multivariate analysis.
- **Multivariate analysis** (**MVA**) is based on the statistical principle of multivariate statistics, which involves observation and analysis of more than one statistical outcome variable at a time. In design and analysis, the technique is used to perform trade studies across multiple dimensions while taking into account the effects of all variables on the responses of interest.

서울아산병원 Asan Medical Center



원 저

아토피척수염의 특징과 개회충척수염과의 연관성

충남대학교 의학전문대학원 충남대학교병원 신경과

이경재 백인철 이주연 김성보 박상민 임중걸 이애영 손은희

Characteristics of Atopic Myelitis and its Relationship With *Toxocara Canis* Myelitis

Kyung Jae Lee, MD, In Chul Baek, MD, Juyoun Lee, MD, Sungbo Kim, MD, Sangmin Park, MD, Jung Geol Lim, MD, Ae Young Lee, MD, Eun Hee Sohn, MD

Department of Neurology, Chungnam National University Hospital, Chungnam National University School of Medicine, Daejeon, Korea

Background: It has recently been reported that atopic disorders are associated with various neurological diseases. Atopic myelitis (AM) has been defined as an idiopathic myelitis with either (1) an atopic disease such as atopic dermatitis, allergic rhinitis, or bronchial asthma, or (2) positive mite antigen-specific immunoglobulin E (IgE) with hyper-IgE-emia. The aim of this study was to characterize the clinical, radiological, and laboratory profiles of AM.
Methods: The characteristics of a group of patients with AM (n=18) were compared with those of a group of non-AM subjects (n=13). The history, clinical symptoms, serologic and cerebrospinal fluid (CSF) findings, and brain and spinal-cord magnetic resonance imaging (MRI) scans of all of the subjects were reviewed.
Results: Clinically, nonacute onset occurred more frequently in the AM group than in the non-AM group (72.2% vs. 30.8%, p=0.033). The positivity of specific IgG antibody to *Toxocara canis* was greater in the AM group than in the non-AM group (p=0.045). Swelling (p=0.001) and gadolinium enhancement (p=0.014) on MRI was also m group. There was tendency toward a greater proportion of males, purely sensory symptoms, and eosinophilia in the AM group than in the non-AM group.
Conclusions: These findings suggest that AM is responsible for a considerable proportion myelitis. It appears that the progression of symptom onset is slower, the positivity of specific canis higher, and the occurrence of swelling and enhancement of the lesion on spinal-cord MRI those with AM than those without.
J Korean Neurol Assoc 31(3):158-164, 2013

Key Words: Myelitis, Atopy, Toxocara, Recurrence

통계는 SPSS 19 통계 프로그램을 이용하였다. 아토피척수염군과 비아토피척수염군 간의 차이를 보기 위해 카이제곱, 피셔(Fisher)의 정확검정과 Mann-Whitney U 검정을 이용하여 비교 분석하였고 아토피척수염에 영향을 미치는 인자를 알아보기 위해 로지스틱 회귀분석을 사용하였다. 통계 유의수준은 p<0.05로 하였다.

서울아산병원 Asan Medical Center

Table 2. Comparison of clinicolaboratory features between patients with and without atopic myelitis

	AM (n=18)	Non-AM (n=13)	*p* value
Mean age of onset (yr)	48.5±11.3	47.9±17.3	ns
Sex (male:female)	17:1	8:5	ns
Known history of atopy (%)	2 (11.1)	0 (0)	ns
Mode of onset			0.033
Acute (%)	5 (27.8)	9 (69.2)	
Non-acute (%)	13 (72.2)	4 (30.8)	
Mean duration of initial symptom to maximal deficit (d)	37.6±27.7	24.8±33.6	ns
Clinical symptoms			
Asymmetry (%)	13 (72.2)	7 (53.8)	ns
Only sensory symptom (%)	15 (83.3)	7 (53.8)	ns
Mean MRC	4.7±0.7	4.2±1.2	ns
Mean MRC sumscore	58.8±2.9	57.5±4.5	ns
Mean maximal EDSS	2.3±1.9	3.1±2.9	ns
Sphincter symptom (%)	10 (55.6)	5 (38.5)	ns
Recurrence (%)	6 (33.3)	2 (15.4)	ns
Duration of follow-up (mon)	30.4±18.2 (range: 12-83)	28.2±24.7 (range: 12-97)	ns
Serologic findings			
Mean eosinophil (%)	4.0±4.8	2.2±2.3	ns
Specific IgG to *Toxocara canis* (%)	10/12 (83.3)	2/7 (28.6)	0.045
Mean *Toxocara canis* titers	0.35±0.11	0.15±0.11	0.002
CSF findings			
Pleocytosis (>5 cells) (%)			
Increased protein (>45 mg/dL) (%)			
OCB (%)			

AM; Atopic myelitis, MRC; Medical Research Council, EDSS; Exp
ns; Not significant.

Table 3. Comparison of radiologic findings between patients with and without atopic myelitis

	AM (n=18)	Non-AM (n=13)	*p* value
Spinal cord MRI findings			
Mean lesion length (vertebral body count)	2.9±1.8	2.9±1.8	ns
Swelling (%)	21/25 (84.0)	4/15 (26.7)	0.001
Gd enhancement (%)	21/23 (91.3)	7/14 (50.0)	0.014
Involved level			
Cervical (%)	5 (20.0)	2 (13.3)	ns
Cervical to thoracic (%)	2 (8.0)	4 (26.7)	ns
Thoracic (%)	18 (72.0)	6 (40.0)	ns
Thoracic to lumbar (%)	0 (0.0)	3 (20.0)	ns
Cervical to lumbar (%)	0 (0.0)	0 (0.0)	ns
Brain MRI findings			
Atypical lesions (%)	6 (35.3)	7 (53.8)	ns
Within normal limit (%)	11 (64.7)	6 (46.2)	ns

AM; Atopic myelitis, MRI; Magnetic resonance imaging, Gd; Gadolinium, ns; Not significant.

Table 4. Logistic regression results for clinicoradiologic factors affecting atopic myelitis

	B	SE	Wald	Df	Sig.	Exp (B)	95% CI for Exp (B)	
							Lower	Upper
Male	2.363	1.176	4.036	1	0.045	10.625	1.059	106.573
Non-acute onset	1.766	0.799	4.891	1	0.027	5.850	1.222	27.994
Toxocara canis (+)	2.526	1.140	4.907	1	0.027	12.500	1.338	116.796
Swelling (MRI)	2.890	0.961	9.045	1	0.003	18.000	2.737	118.391
Enhancement (MRI)	1.946	0.951	4.185	1	0.041	7.000	1.085	45.160

MRI; Magnetic resonance imaging.

Interaction-1 (Wikipedia)

- In general, **Interaction** is a kind of action that occurs as two or more objects have an effect upon one another. The idea of a two-way effect is essential in the concept of interaction, as opposed to a one-way causal effect. A closely related term is interconnectivity, which deals with the interactions of interactions within systems: combinations of many simple interactions can lead to surprising emergent phenomena.

- In statistics, an **interaction** may arise when considering the relationship among three or more variables, and describes a situation in which the simultaneous influence of two variables on a third is not additive. Most commonly, interactions are considered in the context of regression analyses.

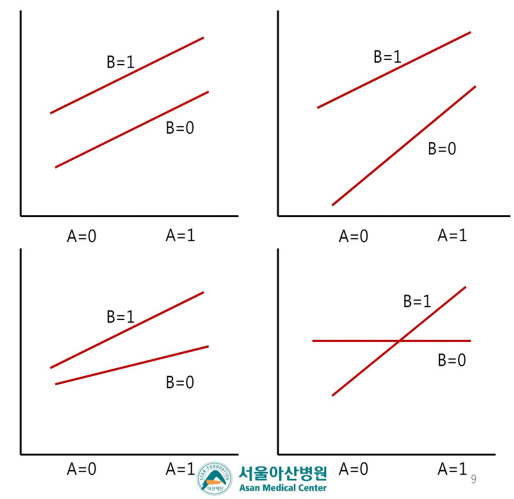
서울아산병원 Asan Medical Center 7

Interaction-2 (Wikipedia)

- The presence of interactions can have important implications for the interpretation of statistical models. If two variables of interest interact, the relationship between each of the interacting variables and a third "dependent variable" depends on the value of the other interacting variable. In practice, this makes it more difficult to predict the consequences of changing the value of a variable, particularly if the variables it interacts with are hard to measure or difficult to control.

- The notion of "interaction" is closely related to that of "moderation" that is common in social and health science research:
- The interaction between an explanatory variable and an environmental variable suggests that the effect of the explanatory variable has been moderated or modified by the environmental variable.

서울아산병원 Asan Medical Center 8



서울아산병원 Asan Medical Center 9

Confounding vs. Interaction

- In clinical trials, confounding and interaction effects are the most common distortions in the evaluation of medication.

- Confounding effects are contributed by various factors such as race & age & gender that cannot be separated by the design under the study.
- Interaction effect between factors is a joint effect with one or more contributing factors(Chow & Liu, 1995)

- Confounding: impossible to assess the treatment effects
- Interaction: possible

서울아산병원 Asan Medical Center 10

Definition of Confounding

- A confounding variable (also confounding factor, lurking variable, a confound, or confounder) is an extraneous variable in a statistical model that correlates (positively or negatively) with both the dependent variable and the independent variable.

- The methodologies of scientific studies therefore need to control for these factors to avoid a false positive (Type I) error; an erroneous conclusion that the dependent variables are in a causal relationship with the independent variable. Such a relation between two observed variables is termed a spurious relationship.

- Thus, confounding is a major threat to the validity of inferences made about cause and effect, i.e. internal validity, as the observed effects should be attributed to the independent variable rather than the confounder.

서울아산병원 Asan Medical Center 11

Introduction By Examples

서울아산병원
Asan Medical Center

ARTICLES

Hippocampal mean diffusivity and memory in healthy elderly individuals

A cross-sectional study

Giovanni A. Carlesimo, MD, PhD
Andrea Cherubini, PhD
Carlo Caltagirone, MD
Gianfranco Spalletta, MD, PhD

Address correspondence and reprint requests to Dr. Giovanni A. Carlesimo, Fondazione IRCCS S. Lucia, V. Ardeatina, 306, 00179 Rome, Italy
memolab@hsantalucia.it

ABSTRACT

Background: The relationship between age-related memory decline and hippocampal anatomic changes is a matter of debate.

Objectives: To investigate the relationship between age-related memory decline and MRI hippocampal anatomic changes in a cohort of healthy individuals.

Methods: In this cross-sectional study, 76 healthy individuals (44 male and 32 female), ranging in age from 20 to 80 years, were recruited from universities, community recreational centers, hospital personnel, and patients' relatives from 2005 to 2008. These individuals were submitted to a 3-T MRI protocol with a whole-brain T1-weighted and diffusion-weighted scanning and a neuropsychological assessment. For each subject, we calculated the volumes of the total brain (gray + white matter) and hippocampi. The segmented hippocampi defined the binary masks where mean values of mean diffusivit… cal evaluation included t… visuospatial memory (20…

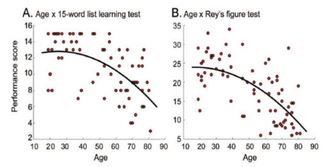
Results: Hippocampal MD… dicted performance of inc…

Conclusions: High mean … uals predict memory de… visual-spatial memory. N…

GLOSSARY

AD = Alzheimer disease; DTI = … interest; TE = echo time; TR = …

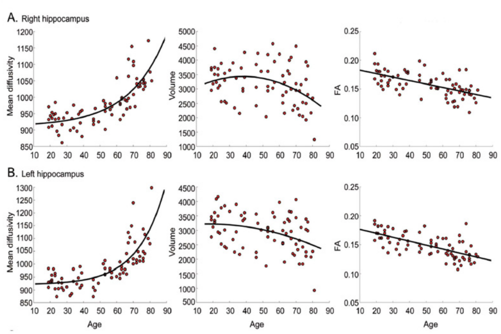
Figure 1 Age and memory performance in healthy individuals



Regression curves that best fit the relationship in the experimental group between age and performance score on the delayed recall of a 15-word list (A) and age and performance score on the delayed reproduction of Rey figure test (B).

Asan Medical Center

Figure 2 Age and anatomic variables of interest in healthy individuals



서울아산병원
Asan Medical Center

Mood, anxiety, and incomplete seizure control affect quality of life after epilepsy surgery

Hamada Hamid, DO, MPH
Karen Blackmon, PhD
Xiangyu Cong, PhD, MPH
James Dziura, PhD, MPH
Lauren Y. Atlas, PhD
Barbara G. Vickrey, MD, MPH
Anne T. Berg, PhD
Carl W. Bazil, MD, PhD
John T. Langfitt, PhD
Thaddeus S. Walczak, MD
Michael R. Sperling, MD
Shlomo Shinnar, MD, PhD
Orrin Devinsky, MD

Correspondence to Dr. Hamid: hamada.hamid@yale.edu

ABSTRACT

Objective: We examined the complex relationship between depression, anxiety, and seizure control and quality of life (QOL) outcomes after epilepsy surgery.

Methods: Seven epilepsy centers enrolled 373 patients and completed a comprehensive diagnostic workup and psychiatric and follow-up QOL evaluation. Subjects were evaluated before surgery and then at 3, 6, 12, 24, 48, and 60 months after surgery. Standardized assessments included the Quality of Life in Epilepsy Inventory-89, Beck Depression Inventory (BDI), and Beck Anxiety Inventory (BAI). A mixed-model repeated-measures analysis was used to analyze associations of depression, anxiety, seizure outcome, and seizure history with overall QOL score and QOL subscores (cognitive distress, physical health, mental health, epilepsy-targeted) prospectively.

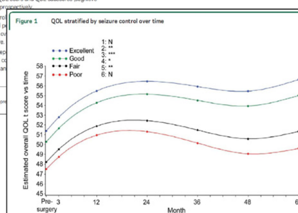
Results: The groups with excellent and good seizure control… on the overall QOL compared to the groups with fair and p… scores were both highly and negatively associated with ov… scores were associated with decreased overall QOL score.

Conclusions: Depression and anxiety are strongly and indep… after epilepsy surgery. Interestingly, even partial seizure c… anxiety levels, improved QOL. Management of mood and an… surgical care. Neurology® 2014;82:887–894

GLOSSARY

AED = antiepileptic drug; BAI = Beck Anxiety Inventory; BDI = Beck Depre… 89 = Quality of Life in Epilepsy Inventory-89.

Figure 1 QOL stratified by seizure control over time



Least squares mean plot based on the modeling results shown in table 2 (5-year seizure-control variable) demonstrates the relationship between quality of life (QOL) and seizure control over time. The y-axis provides overall QOL T scores and color codes the seizure-control status (excellent, good, fair, and poor). The x-axis shows time in months, with the 0 point designated as the presurgical evaluation. Six pairwise comparisons of overall effect of different seizure-control groups over QOL T score: 1 = excellent vs good; 2 = excellent vs fair; 3 = excellent vs poor; 4 = good vs fair; 5 = good vs poor; 6 = fair vs poor. The p value is represented as N = not significant; *p < 0.05; **p < 0.01; *** p < 0.001.

서울아산병원
Asan Medical Center

Table 2 Multivariable repeated-measures regression estimates of the effects of 5-year seizure control, depression, and anxiety on overall QOL and QOL subscores

	Overall QOL		Seizure QOL		Physical QOL		Mental health QOL		Cognitive QOL	
Characteristics	Estimate	p Value	Estimate	p Value	Estimate	p Value	Estimate	p Value	Estimate	p Value
Time	0.45	<0.0001	0.74	<0.0001	0.33	<0.0001	0.25	<0.0001	0.24	<0.0001
Time*time	−0.02	<0.0001	−0.03	<0.0001	−0.01	<0.0001	−0.009	<0.0001	−0.009	<0.0001
Time*time*time	0.0002	<0.0001	0.0002	<0.0001	0.0001	<0.0001	0.0001	0.0001	0.0001	<0.0001
No. of AEDs (full dose)	−0.38	0.12	−0.63	0.02	−0.23	0.43	−0.06	0.80	−0.36	0.25
Seizure control		<0.0001		<0.0001		0.002		0.0002		0.05
Excellent	3.78	<0.0001	3.74	<0.0001	3.57	0.0005	2.89	0.0003	2.95	0.01
Good	0.58	0.6	−0.23	0.85	0.73	0.58	−0.33	0.75	1.87	0.22
Fair	2.69	0.002	2.82	0.003	1.48	0.15	1.99	0.01	3.08	0.01
Poor	Ref		Ref		Ref		Ref		Ref	
Time and seizure control interaction		0.06		<0.0001		0.02		0.52		0.43
Time and excellent	0.05	0.01	0.11	<0.0001	0.05	0.01	−0.01	0.42	0.02	0.33
Time and good	0.02	0.37	0.06	0.009	0.02	0.34	−0.009	0.69	−0.009	0.73
Time and fair	0.04	0.03	0.08	0.0001	0.06	0.004	0.007	0.72	−0.004	0.86
Time and poor	Ref		Ref		Ref		Ref		Ref	
Duration of seizure history, y	0.01	0.5	0.06	0.007	−0.01	0.65	−0.001	0.95	−0.003	0.93
Sex										
Female	0.10	0.85	0.54	0.31	−0.63	0.30	0.67	0.12	−0.38	0.61
Male	Ref		Ref		Ref		Ref		Ref	
Race										
Nonwhite	0.76	0.25	0.48	0.49	1.46	0.06	1.16	0.04	−0.64	0.51
White	Ref		Ref		Ref		Ref		Ref	
Education		0.99		0.14		0.43		0.29		0.51
Graduate school	−0.05	0.95	−1.58	0.06	0.91	0.34	−0.83	0.22	1.27	0.28
College or partial college	−0.07	0.9	−0.75	0.20	0.77	0.24	−0.65	0.17	0.63	0.44
High school or less	Ref		Ref		Ref		Ref		Ref	
Resection location										
Extratemporal	0.09	0.9	0.49	0.55	−0.95	0.30	−0.14	0.83	1.11	0.33
Temporal	Ref		Ref		Ref		Ref		Ref	
Side of brain										
Left	−0.84	0.09	−0.24	0.66	0.22	0.72	0.08	0.84	−2.74	0.0002
Right	Ref		Ref		Ref		Ref		Ref	
BDI score	−0.83	<0.0001	−0.68	<0.0001	−0.63	<0.0001	−0.86	<0.0001	−0.67	<0.0001
Time and BDI interaction	0.003	0.0009	0.004	0.001	0.003	0.01	0.002	0.13	0.004	0.0005
BAI score	−0.44	<0.0001	−0.41	<0.0001	−0.38	<0.0001	−0.48	<0.0001	−0.24	<0.0001
Time and BAI interaction	−0.002	0.03	−0.001	0.29	−0.001	0.24	−0.0009	0.36	−0.004	0.0002
BDI and BAI interaction	0.009	<0.0001	0.008	<0.0001	0.005	0.007	0.009	<0.0001	0.007	<0.0001

Abbreviations: AED = antiepileptic drug; BAI = Beck Anxiety Inventory; BDI = Beck Depression Inventory; QOL = quality of life; Ref = reference.

	Overall QOL		Seizure QOL		Physical QOL		Mental health QOL		Cognitive QOL	
Characteristics	Estimate	p Value	Estimate	p Value	Estimate	p Value	Estimate	p Value	Estimate	p Value
Time	0.45	<0.0001	0.74	<0.0001	0.33	<0.0001	0.25	<0.0001	0.24	<0.0001
Time*time	−0.02	<0.0001	−0.03	<0.0001	−0.01	<0.0001	−0.009	<0.0001	−0.009	<0.0001
Time*time*time	0.0002	<0.0001	0.0002	<0.0001	0.0001	<0.0001	0.0001	0.0001	0.0001	<0.0001
No. of AEDs (full dose)	−0.38	0.12	−0.63	0.02	−0.23	0.43	−0.06	0.80	−0.36	0.25
Seizure control		<0.0001		<0.0001		0.002		0.0002		0.05
Excellent	3.78	<0.0001	3.74	<0.0001	3.57	0.0005	2.89	0.0003	2.95	0.01
Good	0.58	0.6	−0.23	0.85	0.73	0.58	−0.33	0.75	1.87	0.22
Fair	2.69	0.002	2.82	0.003	1.48	0.15	1.99	0.01	3.08	0.01
Poor	Ref		Ref		Ref		Ref		Ref	
Time and seizure control interaction		0.06		<0.0001		0.02		0.52		0.43
Time and excellent	0.05	0.01	0.11	<0.0001	0.05	0.01	−0.01	0.42	0.02	0.33
Time and good	0.02	0.37	0.06	0.009	0.02	0.34	−0.009	0.69	−0.009	0.73
Time and fair	0.04	0.03	0.08	0.0001	0.06	0.004	0.007	0.72	−0.004	0.86
Time and poor	Ref		Ref		Ref		Ref		Ref	

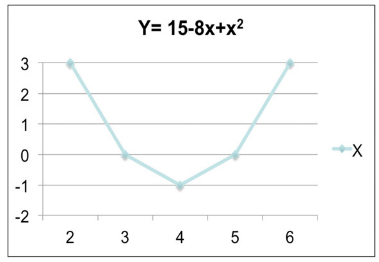
서울아산병원
Asan Medical Center

Goals for today

- A brief refresher of algebra basics
- Modeling simple curvilinear effects in logistic regression
- Graphing curvilinear effects
- Interpreting curvilinear effects
- Curvilinear effects and data cleaning
- Advanced Topic: curvilinear interactions

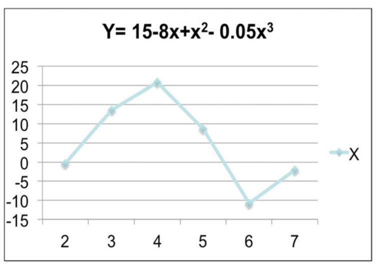
Quick math review

- A simple quadratic equation:
  - $Y = 15 - 8x + x^2$
  - Use x= 2, 3, 4, 5, 6 to sketch out the graph

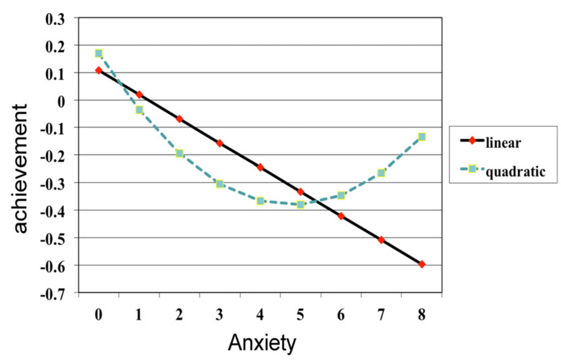


Cubic curves have 2 inflection points

- A simple cubic equation:
  - $Y = 15 - 8x + x^2 - 0.05x^3$



Why curvilinear regression?



Assumption of linearity

- Regression Model
  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$
- In OLS regression we assume that IV is linearly related to DV
- In logistic regression we assume the IV is "linear on the logit"
  - Logit($\hat{Y}$) is linearly related to IV
- When violated: mis-estimation of model

Causes of curvilinearity

- True nonlinear relationship
- Influential data points
- Violating equal intervals in coding of data

Detection of nonlinear effects

- Theory
- Ad hoc testing (quadratic, cubic term)
- Box-Tidwell transformation

서울아산병원 Asan Medical Center

The rules

- test quadratic curve via squaring independent variable
  - Logit($\hat{Y}$) = a + b_1x + b_2x^2
- test cubic curve via cubing IV
  - Logit($\hat{Y}$) = a + b_1x + b_2x^2 + b_3x^3
- Can only test squared term with X in equation
- Can only test cubed term with X and X^2 in equation

서울아산병원 Asan Medical Center

Fitting Curvilinear Model

서울아산병원 Asan Medical Center

Curvilinear logistic regression in practice

Table 7.x

Relationship of Age and Benign forgetfulness –linear model only

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	AGE	.045	.001	1381.895	1	.000	1.046	1.044	1.048
	Constant	-4.631	.074	3890.651	1	.000	.010		

a. Variable(s) entered on step 1: AGE_P.

서울아산병원 Asan Medical Center

Add quadratic term

- Significant Δ-2LL ($X^2_{(1)}$ = 298.45, p < .0001)

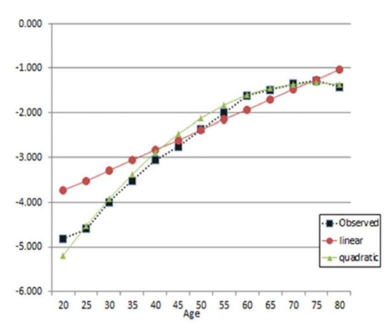
Table 7.x

Predicting benign forgetfulness from Age and Age²

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	AGE	.19402	.010	405.285	1	.000	1.214	1.191	1.237
	AGE²	-.001301	.000	253.400	1	.000	.999	.999	.999
	Constant	-8.56625	.275	971.730	1	.000	.000		

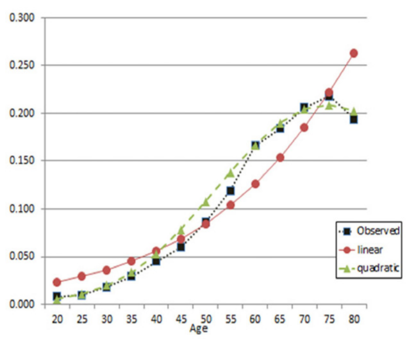
Asan Medical Center

Cubic term significant but very small Δ-2LL



서울아산병원 Asan Medical Center

Graphed as conditional probabilities



서울아산병원 Asan Medical Center

Data cleaning & curvilinear effects

서울아산병원 Asan Medical Center

Marijuana use and Student achievement

- No significant linear relationship between reported marijuana use and student achievement.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	ZBYACH	-.197	.107	3.351	1	.067	.821	.665	1.014
	Constant	-1.164	.103	126.848	1	.000	.312		

a. Variable(s) entered on step 1: ZBYACH.

서울아산병원 Asan Medical Center

MJ and zACH- still no data cleaning

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	ZBYACH	-.177	.120	2.167	1	.141	.838	.662	1.060
	zBYACH2	-.039	.102	.149	1	.699	.961	.787	1.174
	Constant	-1.131	.134	71.266	1	.000	.323		

a. Variable(s) entered on step 1: zBYACH2.

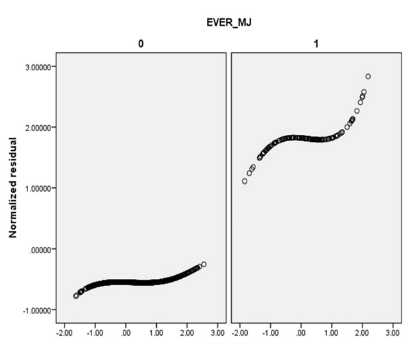
Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	ZBYACH	.055	.212	.067	1	.795	1.057	.697	1.602
	zBYACH2	.078	.134	.340	1	.560	1.081	.831	1.407
	zBYACH3	-.132	.101	1.723	1	.189	.876	.719	1.067
	Constant	-1.199	.145	68.669	1	.000	.302		

a. Variable(s) entered on step 1: zBYACH3.

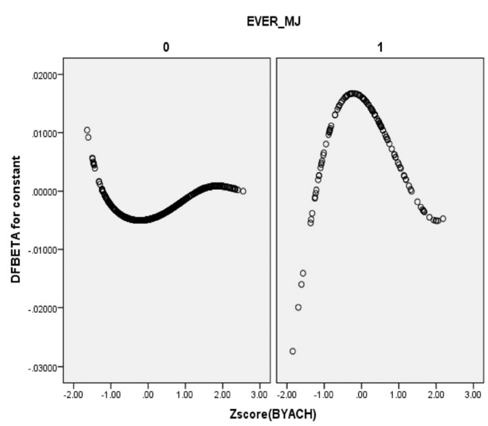
서울아산병원 Asan Medical Center

Standardized residuals



서울아산병원 Asan Medical Center

DfBetas for intercept



After very conservative data cleaning

Regression line equation:

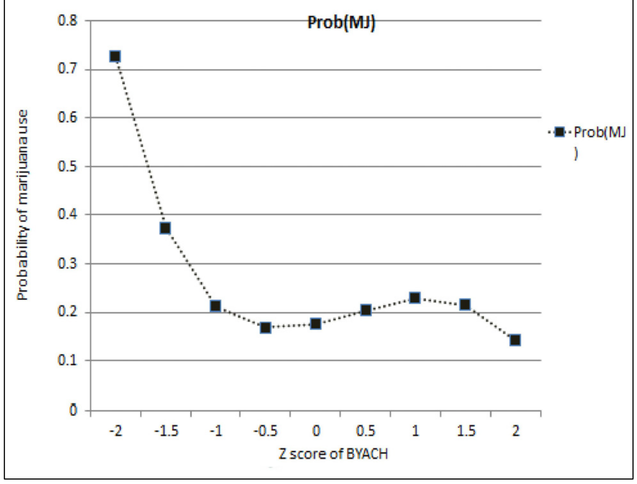
$$\text{Logit}(\hat{Y}) = -1.546 + 0.296(\text{ZBYACH}) +$$

Variables in the Equation

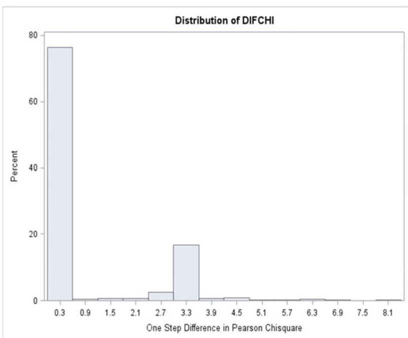
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	ZBYACH	.296	.227	1.695	1	.193	1.344	.861	2.097
	zBYACH2	.285	.148	3.682	1	.055	1.329	.994	1.778
	zBYACH3	-.247	.110	5.075	1	.024	.781	.630	.968
	Constant	-1.546	.166	87.138	1	.000	.213		

a. Variable(s) entered on step 1: zBYACH3.

서울아산병원 Asan Medical Center



Data cleaning via DIFCHISQ



서울아산병원 Asan Medical Center

Comparing DfBeta vs. DIFCHISQ

- After DfBeta cleaning

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	ZBYACH	.296	.227	1.695	1	.193	1.344	.861	2.097
	zBYACH2	.285	.148	3.682	1	.055	1.329	.994	1.778
	zBYACH3	-.247	.110	5.075	1	.024	.781	.630	.968
	Constant	-1.546	.166	87.138	1	.000	.213		

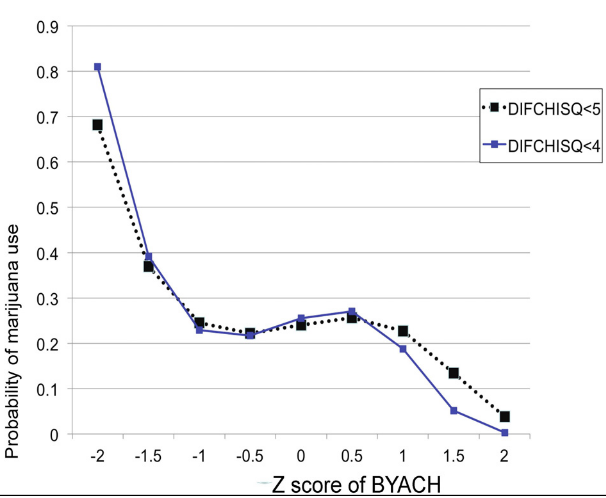
a. Variable(s) entered

After DIFCHISQ cleaning:

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.1514	0.1449	63.1217	<.0001
ZBYACH	1	0.2683	0.2421	1.2283	0.2677
zBYACH2	1	-0.0214	0.1401	0.0234	0.8784
zBYACH3	1	-0.3168	0.1314	5.8118	0.0159

Asan Medical Center



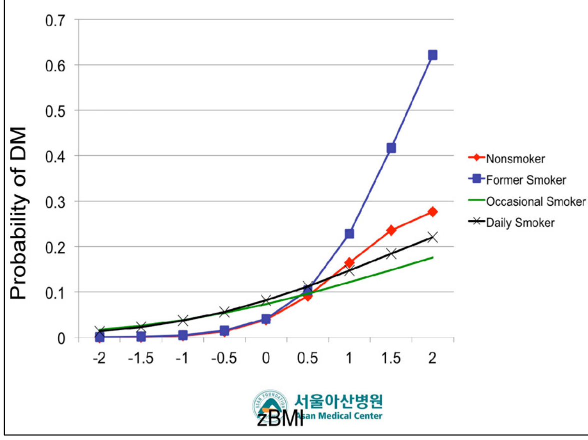
Summary so far...

- Surprisingly simple to find curvilinear effects
- Surprisingly challenging to find examples of data cleaning eliminating curvilinear effects
- This type of analysis is procedurally similar to logistic regression with multiple IVs but all different aspects of SAME IV.

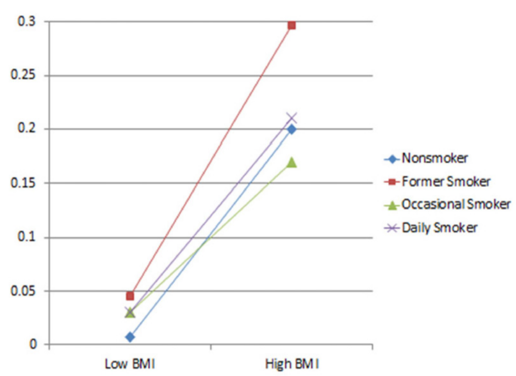
서울아산병원 Asan Medical Center

With Interaction

Complex interactions



BORING!!!



Linear interactions

- An interaction allows for effect of X_1 to be non-constant across the range of X_2

$$\mathrm{Logit}(Y) = b_0 + b_1X_1 + b_2X_2 + b_3X_1X_2$$

- If you as researcher do NOT model interaction, you are explicitly asserting b_3 is 0.
- Should be tested rather than asserted

"Simple" curvilinear effects

- A curvilinear effect allows for modeling of nonlinear relationships

$$\mathrm{Logit}(Y) = b_0 + b_1X_1 + b_2X_1^2 + b_3X_1^3 + \ldots + b_kX_1^k$$

- If you as researcher do NOT model a curvilinear effect, you are explicitly asserting that $b_2 \rightarrow b_k$ all 0.
- Should be tested rather than asserted

Review of procedures

- Z score continuous variables (or center)
- Create powered terms by raising X to a power
  - Such as X^2 and X^3
- Dichotomous and categorical variables cannot have curvilinear effects

- Enter linear effects on a block
- Enter all higher-order terms on another block
- Change in -2LL evaluates whether ANY curvilinear effect is significant

Procedurally...

- **Curvilinear interaction** means that the nature or shape of a curvilinear effect depends upon another variable.
  - i.e. different groups have different curves
- Create the following terms:
  - X and Z are independent variables
  - X^2 and Z^2 are quadratic effects of X and Z
- Create cross-products of these terms
  - XZ
  - X^2Z and XZ^2
  - X^2Z^2

서울아산병원 Asan Medical Center

Two variables, one curvilinear effect

$$\text{Logit}(Y) = b_0 + b_1X + b_2X^2 + b_3Z + b_4XZ + b_5X^2Z$$

or

$$\text{Logit}(Y) = b_0 + b_1X + b_2X^2 + b_3X^3 + b_4Z + b_5XZ + b_6X^2Z + b_7X^3Z$$

서울아산병원 Asan Medical Center

Two curvilinear effects

$$\text{Logit}(Y) = b_0 + b_1X + b_2X^2 + b_3Z + b_4Z^2 + b_5XZ + b_6X^2Z + b_7XZ^2 + b_8X^2Z^2$$

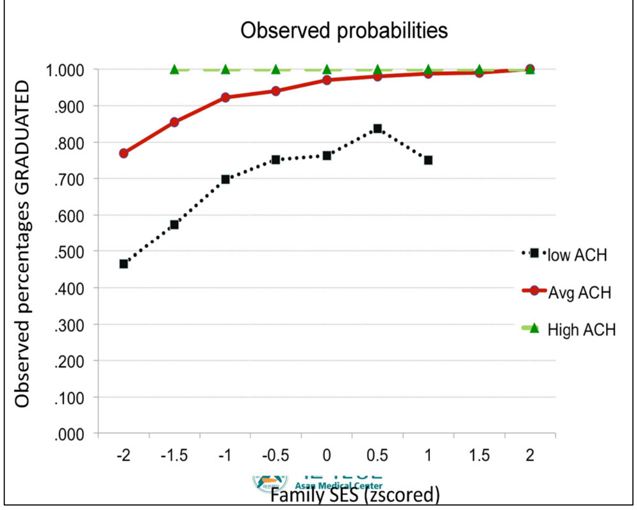
or

$$\text{Logit}(Y) = b_0 + b_1X + b_2X^2 + b_3X^3 + b_4Z + b_5Z^2 + b_6Z^3 + b_7XZ + b_8X^2Z + b_9XZ^2 + b_{10}X^2Z^2 + b_{11}X^3Z + b_{12}X^3Z^2 + b_{13}XZ^3 + b_{14}X^2Z^3 + b_{15}X^3Z^3$$

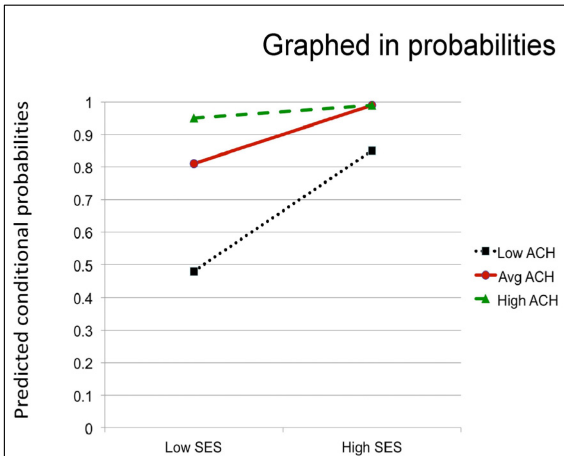
서울아산병원 Asan Medical Center

Prediction

서울아산병원 Asan Medical Center



서울아산병원 Asan Medical Center



서울아산병원 Asan Medical Center

Table 14
Summary of curvilinear interaction model after data cleaning

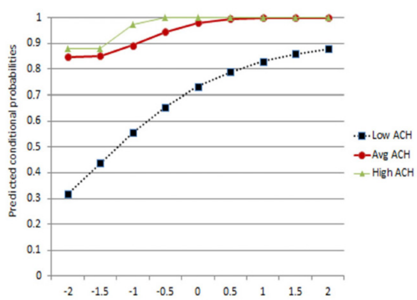
Model	-2LL	Δ -2LL	$p <$ for Δ -2LL
Intercept only	9486.68	--	--
zSES, zACH	7256.97	2229.71	.0001
zSES x zACH	7163.50	93.47	.0001
$zSES^2$, $zACH^2$	7153.59	9.92	.007
$zSES^2$ x zACH, zSES x $zACH^2$, $zSES^2$ x $zACH^2$	7097.97	55.62	.0001

Table 15
Final step of curvilinear interaction analysis after data cleaning ZRE < 5

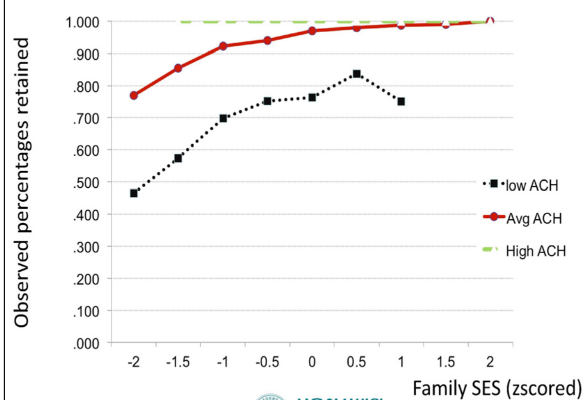
	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
zACH	2.704	.251	115.612	1	.000	14.933	9.123	24.444
zSES	2.451	.226	117.926	1	.000	11.603	7.455	18.060
zACH by zSES	2.709	.435	38.778	1	.000	15.012	6.400	35.215
zACH2	.633	.148	18.236	1	.000	1.883	1.408	2.517
zSES2	.685	.111	37.860	1	.000	1.984	1.595	2.468
zACH by zSES2	.851	.195	19.160	1	.000	2.343	1.600	3.431
zSES by zACH2	.913	.235	15.104	1	.000	2.493	1.573	3.952
zSES2 by zACH2	.229	.108	4.539	1	.033	1.258	1.019	1.553
Constant	3.882	.121	1029.937	1	.000	48.520		

$$\text{Logit}(Y) = 3.882 + 2.704(zACH) + 2.451(zSES) + 2.709(zACH*zSES) + 0.633(zACH^2) + 0.685(zSES^2) + 0.851(zACH * zSES^2) + 0.913(zSES * zACH^2) + 0.229(zSES^2 * zACH^2)$$

Graph of curvilinear interaction



Which is a lot closer to the Observed probabilities



Composing all cross-products for a curvilinear interaction equation

	X	X^2	X^3
Dum1	X Dum1	X^2 Dum1	X^3 Dum1
Dum2	X Dum2	X^2 Dum2	X^3 Dum2
Dum3	X Dum3	X^2 Dum3	X^3 Dum3

Using same DIFCHISQ cleaned data as above:

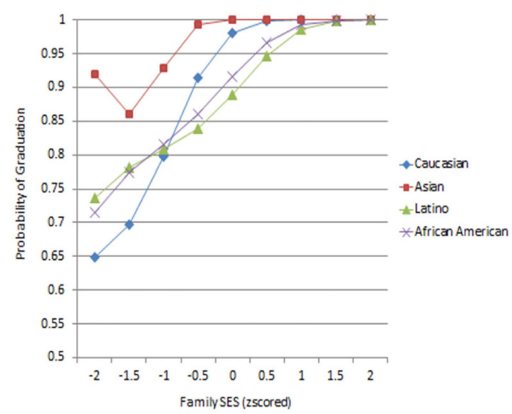
Model	-2LL	Δ -2LL	$p <$ for Δ -2LL
Intercept only	8926.86	--	--
Dum1, Dum2, Dum3	8540.81	386.05	.0001
zSES, $zSES^2$, $zSES^3$	6950.01	1590.80	.0001
zSESxDUM1, zSESxDUM2, zSESxDUM3	6835.11	114.90	.0001
$zSES^2$xDUM1, $zSES^2$xDUM2, $zSES^2$xDUM3	6780.46	54.65	.0001
$zSES^3$xDUM1, $zSES^3$xDUM2, $zSES^3$xDUM3	6774.65	5.81	.12

Table 18
Summary of last significant step in model

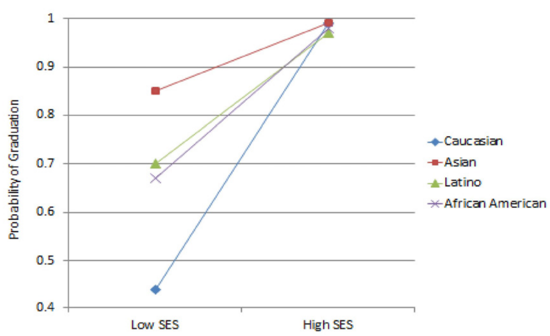
	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
dum1	4.916	2.080	5.583	1	.018	136.399	2.312	8046.207
dum2	-1.868	.161	134.484	1	.000	.154	.113	.212
dum3	-1.556	.177	76.988	1	.000	.211	.149	.299
ZSES	3.907	.250	245.122	1	.000	49.764	30.513	81.161
zses2	1.522	.190	64.447	1	.000	4.582	3.160	6.643
zses3	.202	.049	16.653	1	.000	1.223	1.110	1.348
ZSES by dum1	5.872	3.300	3.166	1	.075	355.094	.551	228788.970
ZSES by dum2	-2.730	.288	89.795	1	.000	.065	.037	.115
ZSES by dum3	-2.410	.327	54.305	1	.000	.090	.047	.170
zses2 by dum1	2.165	1.268	2.917	1	.088	8.718	.726	104.635
zses2 by dum2	-.793	.134	35.020	1	.000	.453	.348	.588
zses2 by dum3	-.738	.152	23.496	1	.000	.478	.355	.644
Constant	3.954	.111	1267.285	1	.000	52.121		

서울아산병원 Asan Medical Center

Result



Original linear interaction



Summary

- Univariate & Multivariate Analysis
- Linear & Curvilinear
- Interactions

- Curves are common if you look for them
- Curvilinear effects often more accurate models of data
- Not difficult
- Careful graphing explicates effects
- Data cleaning always important

서울아산병원 Asan Medical Center