



한양대학교간호과학연구소
Research Institute of Nursing Science Hanyang University

2022

여름방학 통계특강

HANYANG UNIVERSITY

실습을 중심으로 배우는
알기쉬운 SPSS 중급통계



2022년 6월 30일 (목)

CONTENTS

1. 상관관계 분석

(김미영 교수)

- 연구 목적과의 일관성
- 상관분석
- Pearson 상관계수
- SPSS 실습

2. 회귀분석을 위한 가정 검정

(오은미 교수)

- 회귀분석
- 다중회귀분석
- 회귀분석 가정검증
 - 분석전 점검
 - 잔차 분석을 통한 가정진단
- SPSS 실습

3. 선형회귀분석

(김정아 교수)

- 사전작업
 - Data Set 검토
 - Dummy Variable 만들기
- 단순회귀분석 ---- 다중회귀분석----- 위계적 회귀분석
 - 기본 개념과 연구가설
 - SPSS로 단순회귀분석하기

4. 로지스틱 회귀분석

(문설화 교수)

- Logistic regression 소개
- DATA 이용한 SPSS실습
 - Data 확인
 - Univariable analysis
 - Multivariable analysis



01

상관관계 분석

김미영 교수
한양대학교 간호학부

상관관계 분석

측정 자료의 종류

<p>연속형 (continuous)</p>	<p>측정 가능하거나 셀 수 있는 것에 대한 자료(변수)</p> <p>➔ 구간척도(interval)</p> <ul style="list-style-type: none">- 크기의 차이에 의해 상대적 비교가 가능한 경우- (예) 지능지수, 온도 <p>➔ 비척도(ratio)</p> <ul style="list-style-type: none">- 값들의 비가 의미를 갖는 경우로 대부분의 측정형 변수- (예) 키, 몸무게, 나이
<p>범주형 (categorical)</p>	<p>조사대상의 특성을 분류하는데 사용되는 자료(변수)</p> <p>➔ 순위척도(ordinal)</p> <ul style="list-style-type: none">- 크기에 순서가 있는 분류- (예) 병의 중증도 단계, 사회경제수준 <p>➔ 명목척도(nominal)</p> <ul style="list-style-type: none">- 크기는 없이 단지 분류만- (예) 성별, 인종, 병의 유무, 병의 종류

분석을 어떻게 할 것인가?

- 목적에 맞는 적절한 통계분석방법 선택
 - 연구목적과의 일관성
- 연구 목적, 통계분석방법, 결과의 해석, 논의, 결론이 일관성이 있어야함

연구목적과의 일관성

2. 연구의 목적

본 연구의 목적은 00를 대상으로 A, B 및 C의 관계를 알아보기 위함이며, 구체적인 목적은 다음과 같다.

- 1) 00가 지각한 A, B 및 C의 정도를 파악한다.
- 2) 00의 일반적 특성에 따른 A, B 및 C의 차이를 파악한다.
- 3) 00가 지각한 A, B 및 C의 관계를 파악한다.

@ 두 집단의 평균비교

- 독립된 두 집단의 평균비교

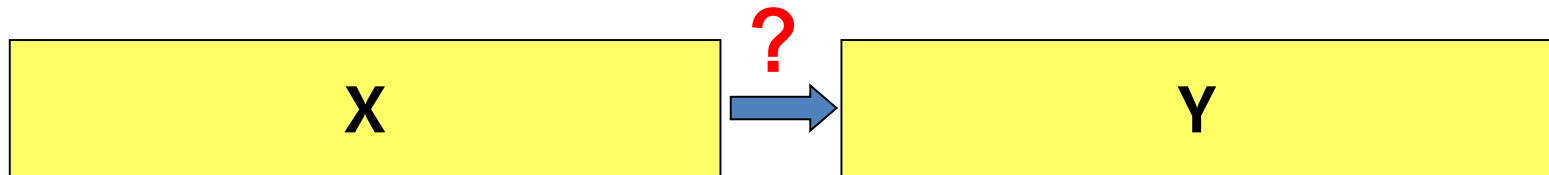
(Independent t-test)

- 짝지어진 자료의 평균비교

(Paired t-test)

@ 세 집단이상의 평균비교(ANOVA)

Correlation vs. Regression



- ⊙ X와 Y 사이 선후관계 불분명하여 상호 영향을 미칠 수 있는 경우 → 상관분석
- ⊙ X의 변동이 Y의 변동에 비해서 시간적으로 앞선다, Y의 변동이 X의 변동에는 영향을 미치지 못한다 → 회귀분석

연구목적과의 일관성

2. 연구의 목적

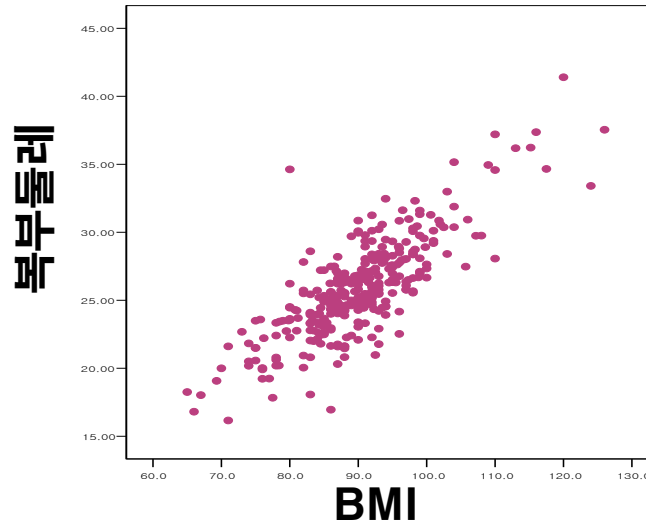
본 연구는 급성기 뇌졸중 생존자를 돌보는 가족의 우울, 돌봄난이도, 생존자의 기억 및 행동 문제에 대한 인식 및 자가 관리의 **정도를 조사하고, 자가 관리에 영향을 미치는 요인을 파악하는 데 목적**이 있다.

상관분석

- 범주형 변수에 대하여 Pearson 상관분석을 수행
- 분석결과를 인과관계인 것처럼 확대 해석하는 것에 주의
- 상관분석의 결과를 인과관계(causality)로 설명하는 경우가 많이 있음.

상관분석

- BMI가 1kg/m^2 증가할 때, 복부둘레는 일정하게 증가하는가?

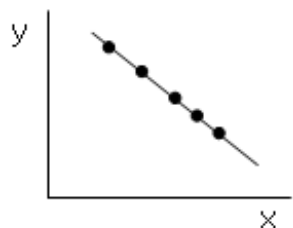


- 변수들 간에 선형적인 관련성이 있는가?
- 선형적 관련성: x 가 1 단위 증가할 때, y 가 일정하게 증가 혹은 감소하는 관련성

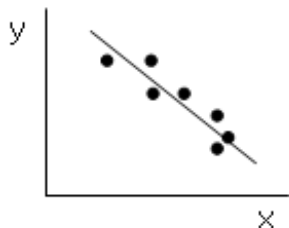
상관분석

- 상관분석이란 두 연속형 변수 x, y 의 선형적 관련성의 정도를 추정하고 검정하는 분석
- 두 연속형 변수에 대한 산점도로 관련성 파악
- 선형적 관련성 정도를 추정
 - 모수적 상관분석: Pearson 상관계수
 - 비모수적 상관분석: Spearman 상관계수,
Kendall's tau-b

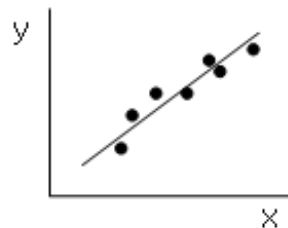
Pearson 상관계수



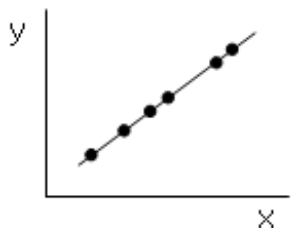
(a) $r = -1$



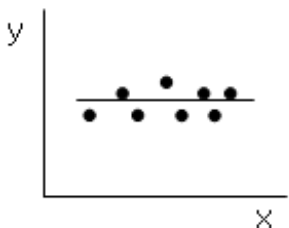
(b) $-1 < r < 0$



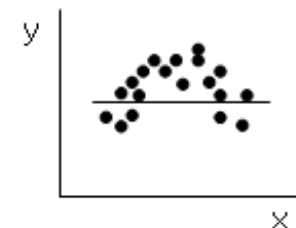
(c) $0 < r < 1$



(d) $r = 1$



(e) $r = 0$



(f) $r = 0$

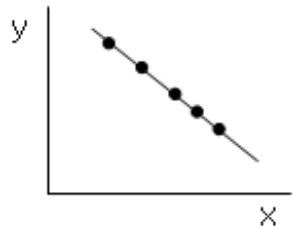
(a) x, y가 정확한 선형 관계:
 $r = -1$

(b) x가 클 때, y는 작아지면:
 $-1 < r < 0$

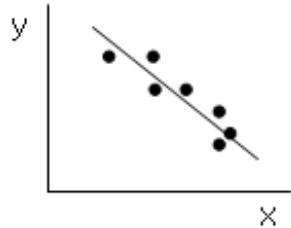
(c) x가 클 때, y도 커지면:
 $0 < r < 1$

(d) x, y가 정확한 선형 관계:
 $r = 1$

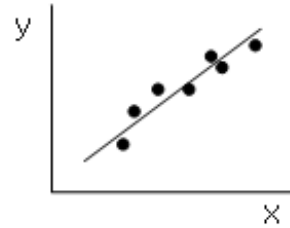
Pearson 상관계수



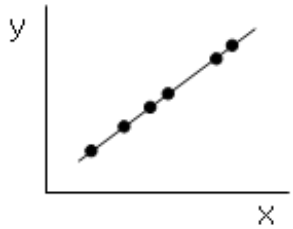
(a) $r = -1$



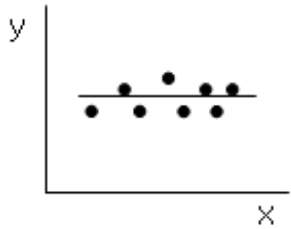
(b) $-1 < r < 0$



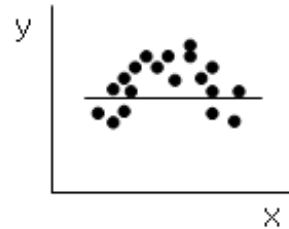
(c) $0 < r < 1$



(d) $r = 1$



(e) $r = 0$



(f) $r = 0$

(e) x, y사이의 어떠한 관련성도 없어 보임.

$$r = 0$$

(f) x가 증가함에 따라 y는 증가 후 감소하는 관련성이 보임. 하지만, 선형적 관련성은 없음.

$$r = 0$$

-> 상관관계가 없다($r=0$)는 것은 두 변수 사이의 관련성이 없음을 의미하지 않고, '선형적 관련성'이 없다는 것을 의미!

Pearson 상관계수의 검정

- Pearson 상관계수의 통계적 검정
(상관계수가 0과 유의하게 다른지 검정)
귀무가설: x, y 사이의 상관관계 없음($r=0$).
대립가설: x, y 사이의 상관관계 있음($r \neq 0$).
- 두 연속 변수(x, y) 중 적어도 하나는 정규분포를 따른다고 가정
- 상관관계는 두 변수의 선형적인 관련성의 정도를 나타내는 것이고, 인과관계를 의미하는 것은 아님(인과관계는 회귀분석).

SPSS 실습

예제: A 병원에서 건강검진을 받은 사람을 대상으로
나이, 성별, 키, 몸무게, 복부둘레등을 측정(N=322)

실습자료: '상관분석 실습자료.sav'

✓ 복부둘레가 증가할 수록 BMI가 증가하는가?

(복부둘레와 BMI의 선형적인 관련성)

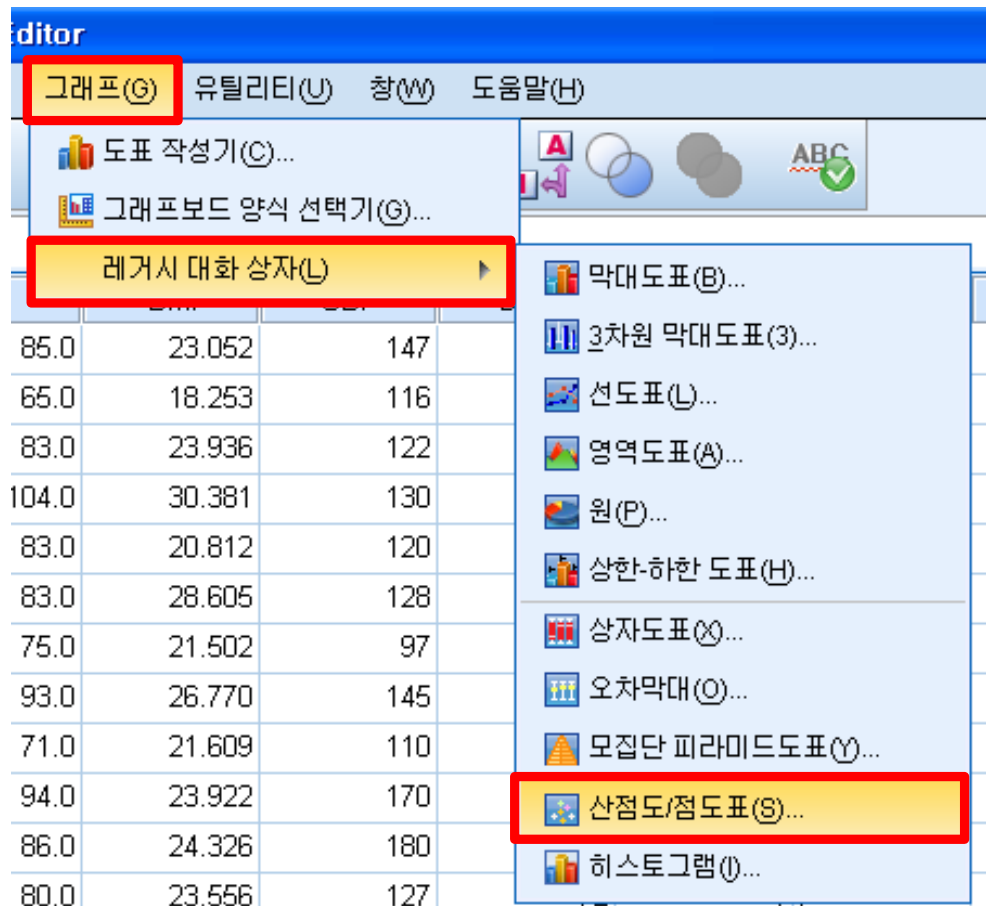
단, 복부둘레와 BMI는 정규분포를 따른다고 가정

- 산점도: 관련성 패턴 파악, 선형적 관련성 여부
- Pearson 상관계수: 관련성의 정도 추정 및 검정

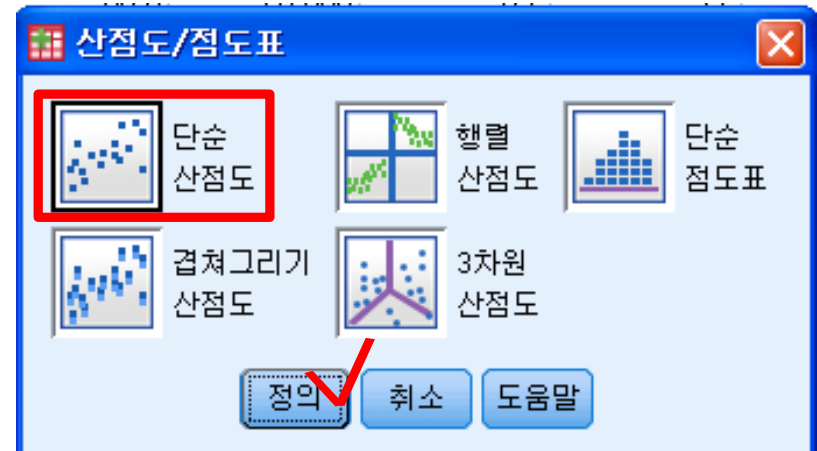
SPSS 실습: 산점도

산점도: 관련성 패턴 파악, 선형적 관련성 여부

그래프 >> 레거시 대화상자 >> 산점도/점도표
(Graphs) (Legacy Dialog) (Scatter/Dot)

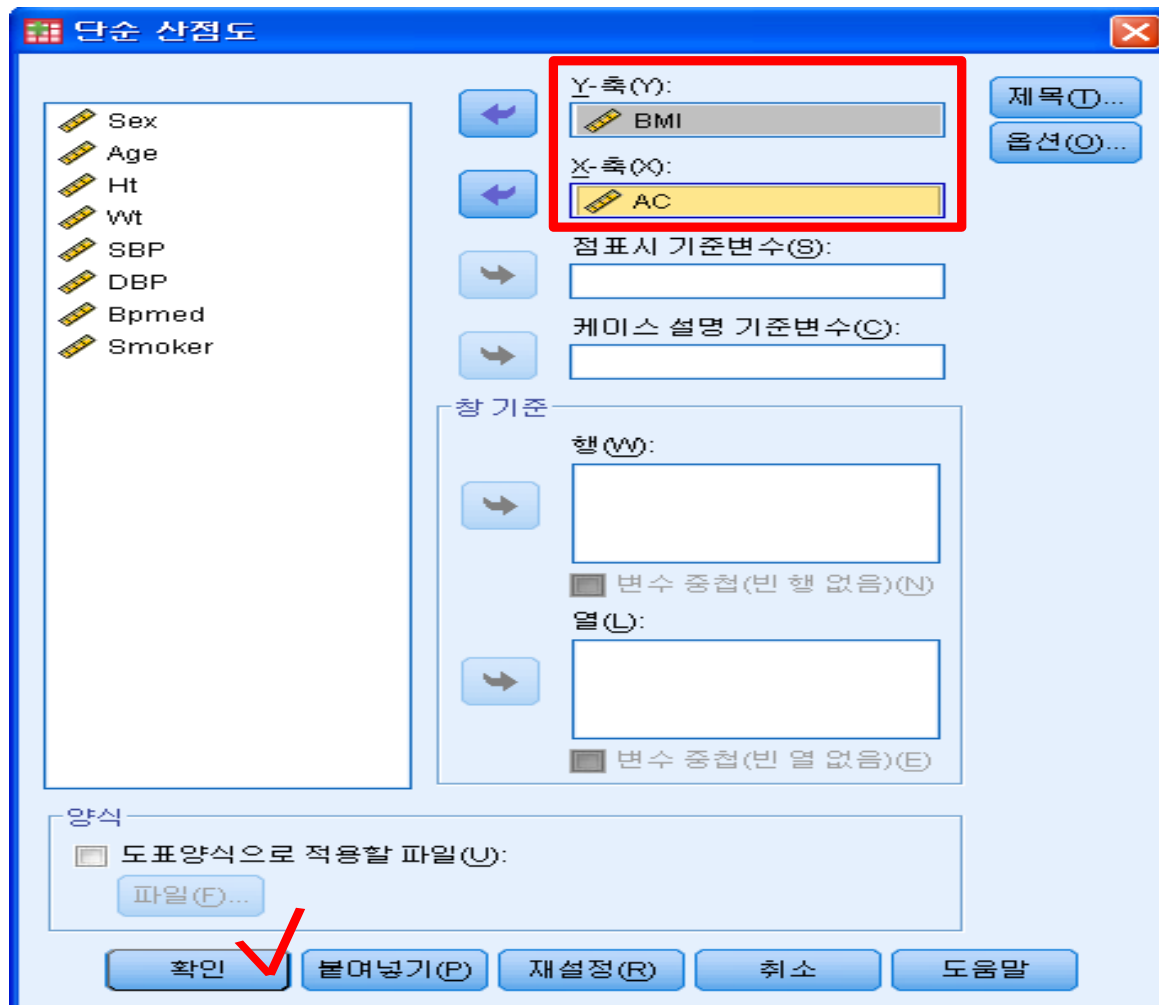


✓ 허리둘레와 BMI의 산점도
: 단순 산점도



SPSS 실습: 산점도

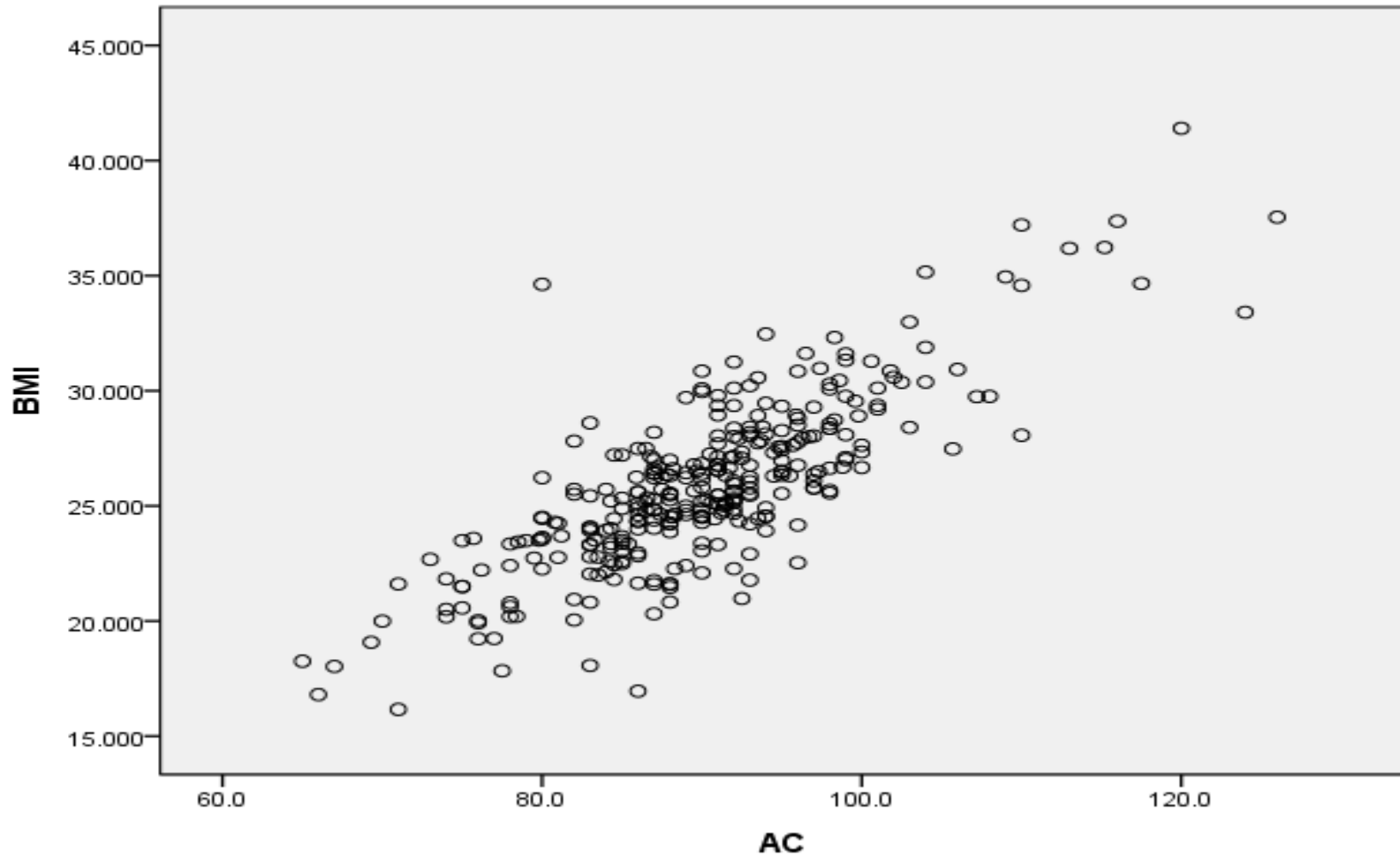
그래프 >> 레거시 대화상자 >> 산점도/점도표 >> 단순산점도
(Graphs) (Legacy Dialog) (Scatter/Dot)



✓ 허리둘레와 BMI의 산점도

Y-축(Y): BMI
X-축(X): AC

SPSS 결과: 산점도

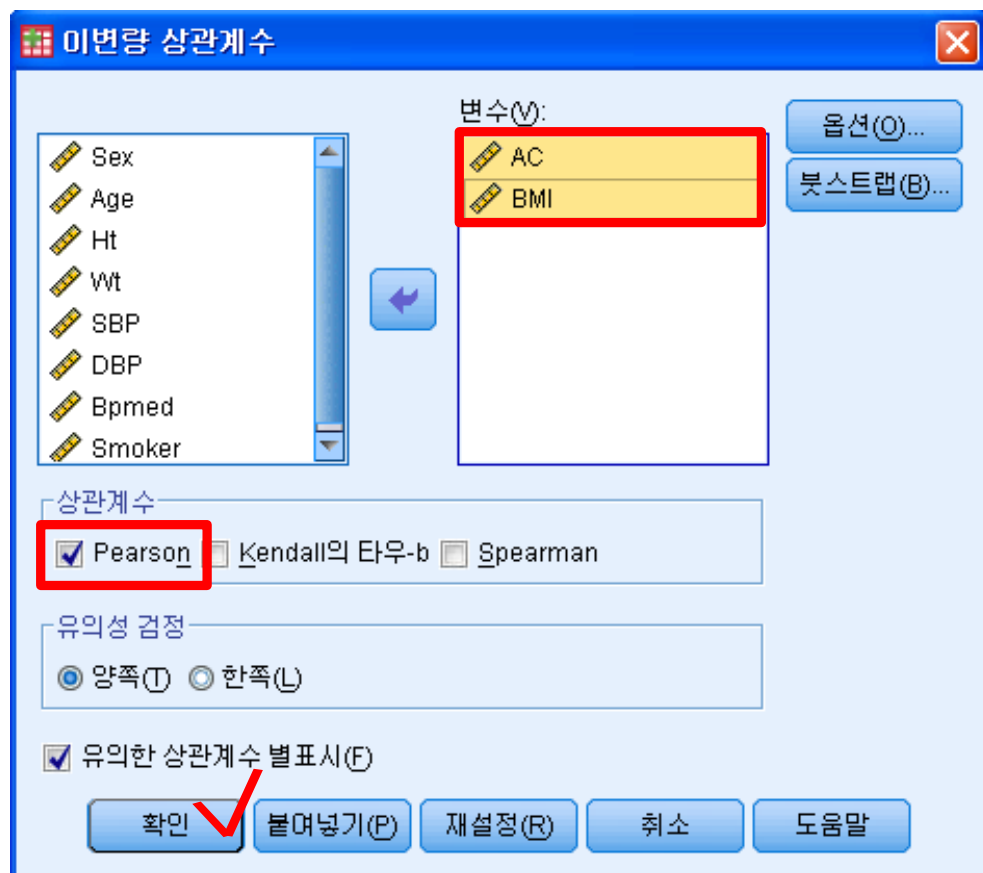
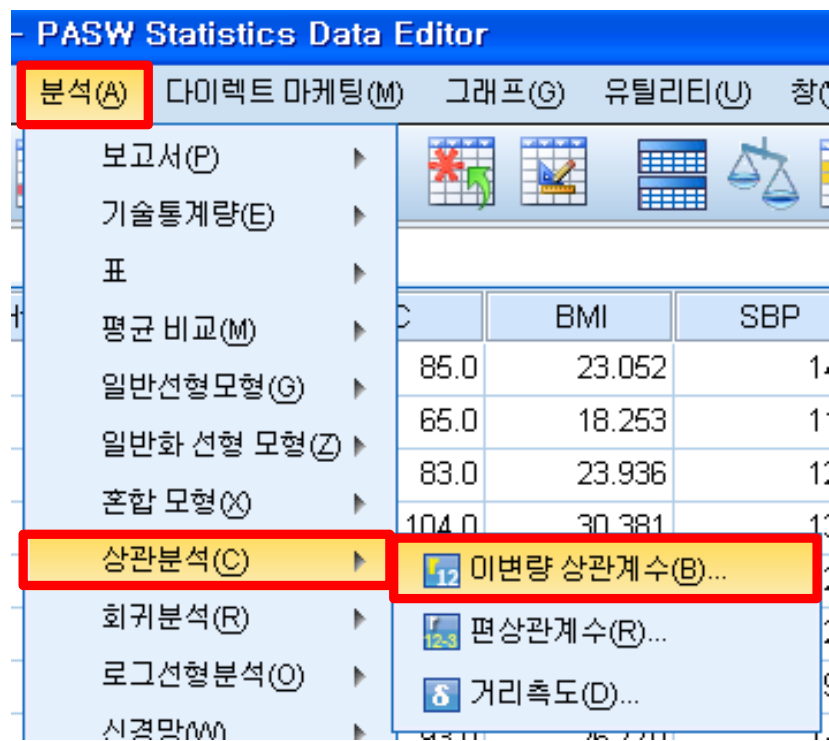


- 복부둘레(AC)가 증가할수록 BMI도 증가하는 경향이 보임.

SPSS 실습: Pearson 상관계수

Pearson 상관계수: 관련성의 정도 추정 및 검정

분석 >> 상관분석 >> 이변량 상관계수
(Analyze) (Correlate) (Bivariate)



SPSS 결과: Pearson 상관계수

상관계수

		AC	BMI
AC	Pearson 상관계수	1	.800**
	유의확률 (양쪽)		.000
	N	322	322
BMI	Pearson 상관계수	.800**	1
	유의확률 (양쪽)	.000	
	N	322	322

** 상관계수는 0.01 수준(양쪽)에서 유의합니다.

- Pearson 상관계수는 0.8로 BMI와 복부둘레(AC)는 유의한 양의 상관관계가 있음.

비모수 상관분석

- 정규분포를 따르지 않는 두 연속변수 혹은 범주형 자료 사이의 관련성의 정도를 추정, 검정 -> Spearman, Kendall's tau

Spearman 상관계수

- 두 변수 x, y 에 대한 순위변수 R_x, R_y 생성

x	y	R_x	R_y
15	116		
20	130		
25	120		
30	132		
40	160		

- 순위변수 R_x, R_y 에 대한 Pearson 상관계수 계산

비모수 상관분석

- 정규분포를 따르지 않는 두 연속변수 혹은 범주형 자료 사이의 관련성의 정도를 추정, 검정 -> Spearman, Kendall's tau

Spearman 상관계수

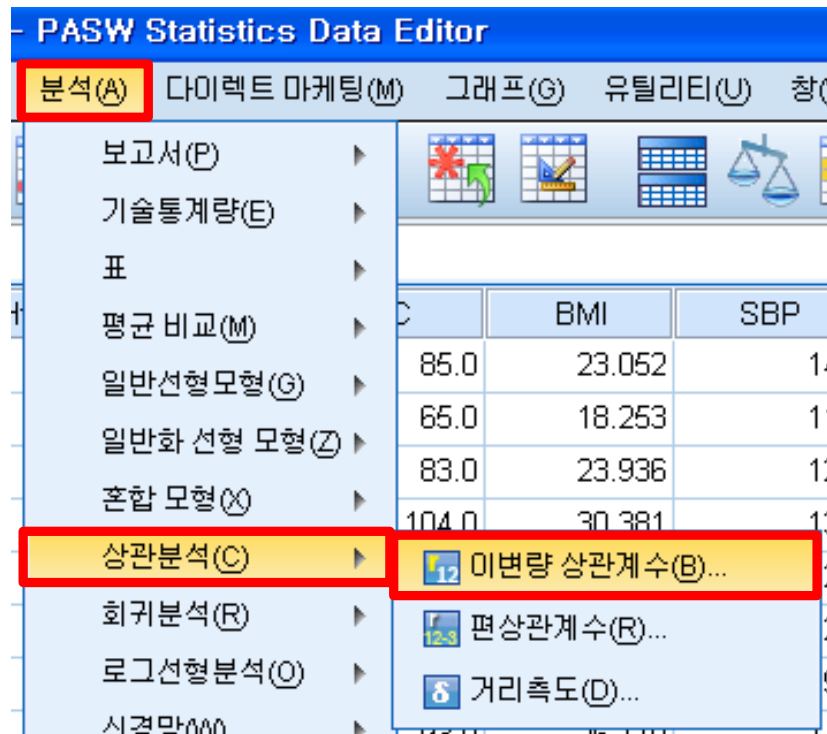
- 두 변수 x, y 에 대한 순위변수 R_x, R_y 생성

x	y	R_x	R_y
15	116	1	1
20	130	2	3
25	120	3	2
30	132	4	4
40	160	5	5

- 순위변수 R_x, R_y 에 대한 Pearson 상관계수 계산

SPSS 실습: Spearman 상관계수

분석 >> 상관분석 >> 이변량 상관계수
(Analyze) (Correlate) (Bivariate)



SPSS 결과: Spearman 상관계수

상관계수			AC	BMI
Kendall의 tau_b	AC	상관계수	1.000	.574**
		유의확률(양측)	.	.000
		N	322	322
	BMI	상관계수	.574**	1.000
		유의확률(양측)	.000	.
		N	322	322
Spearman의 rho	AC	상관계수	1.000	.752**
		유의확률(양측)	.	.000
		N	322	322
	BMI	상관계수	.752**	1.000
		유의확률(양측)	.000	.
		N	322	322

** . 상관 유의수준이 0.01입니다(양측).

- 비모수 상관계수로부터 BMI와 복부둘레(AC)는 유의한 양의 상관관계가 있음.

Table 8. 긍정심리자본, 낙인지각, 사회적지지, 부담감의 상관관계

(N=103)

Variable	부담감	긍정심리자본	낙인지각
	$r(p)$		
긍정심리자본	-.208 (.035)		
낙인지각	.572 (<.001)	-.399 (<.001)	
사회적지지	-.226 (.022)	.595 (<.001)	-.206 (.037)

자가 관리를 잘 하는 경우 우울이 낮게 나타났고,
($r = -.45$, $p < .001$)

자가 관리를 잘 할 수록 우울이 낮게 나타났고,
($r = -.45$, $p < .001$)

- 자가 관리는 우울($r = -.45$, $p < .001$)과 통계적으로 유의한 역상관관계를 나타냈다.

SPSS 실습

감사합니다



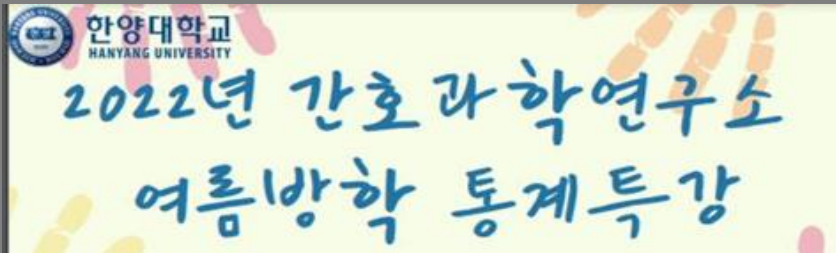
02

회귀분석을 위한 가정 검정

오은미 교수

한양대학교 간호과학연구소

실습을 중심으로 배우는 알기 쉬운 SPSS 중급통계



회귀분석을 위한 가정 검정

2022. 6. 30(목) 10:00-11:40

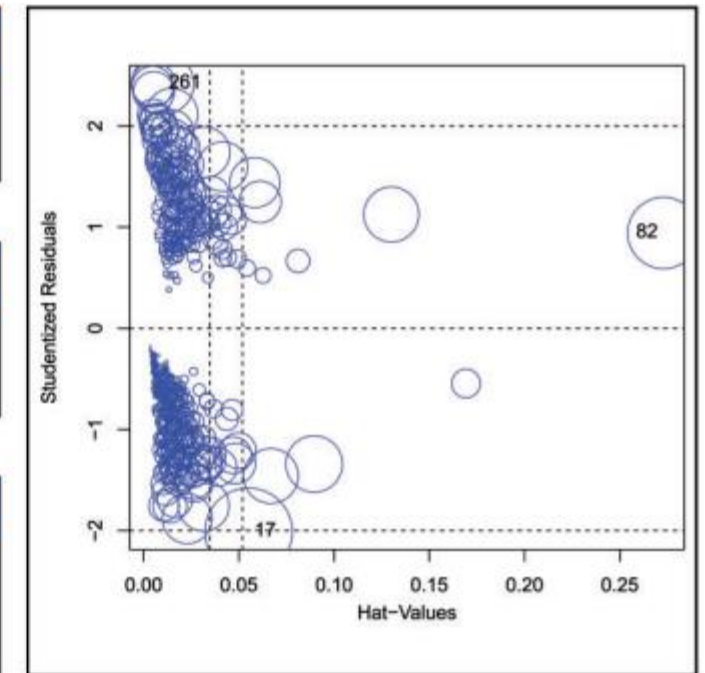
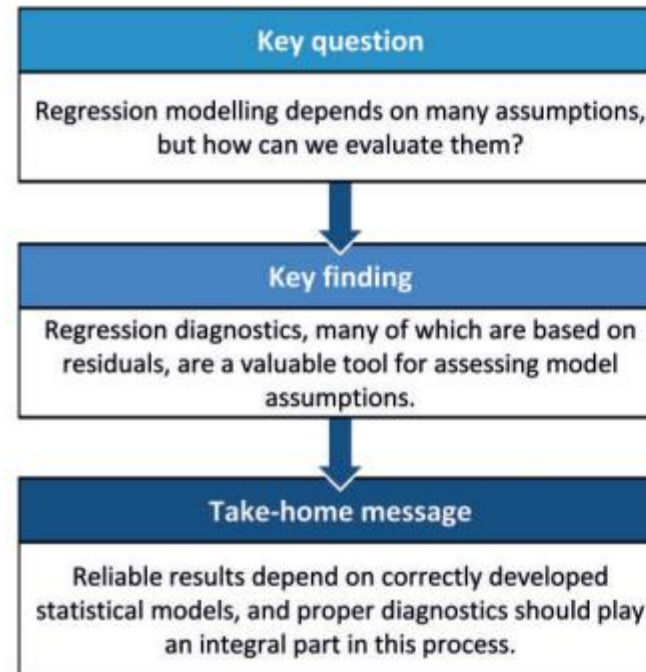
오은미 RN GNP PhD

한양대학교 간호과학연구소 연구교수

Cite this article as: Hickey GL, Kontopantelis E, Takkenberg JJM, Beyersdorf F. Statistical primer: checking model assumptions with regression diagnostics. Interact CardioVasc Thorac Surg 2019;28:1–8.

Statistical primer: checking model assumptions with regression diagnostics[†]

Graeme L. Hickey^{a,*}, Evangelos Kontopantelis^{b,c}, Johanna J.M. Takkenberg^d and Friedhelm Beyersdorf^e



Summary

Regression modelling is an important statistical tool frequently utilized by cardiothoracic surgeons. However, these models—including linear, logistic and Cox proportional hazards regression—rely on certain assumptions. If these assumptions are violated, then a very cautious interpretation of the fitted model should be taken. Here, we discuss several assumptions and report diagnostics that can be used to detect departures from these assumptions. Most of the diagnostics discussed are based on residuals: a measure of the difference between the observed and model fitted values. Reliable and generalizable results depend on correctly developed statistical models, and proper diagnostics should play an integral part in the model development.

Keywords: Statistics • Regression • Residuals • Assumptions • Diagnostics

회귀분석

회귀분석

종속변수와 독립변수 metric variable

선형회귀분석

Linear regression

독립변수가 하나

단일회귀분석

Simple linear regression

독립변수가 두개이상

다중회귀분석

Multiple linear regression

회귀분석

- ▶ 하나 혹은 다수의 독립변수와 한 개의 종속변수 사이의 관계를 검정
 - 단순선형 회귀분석 (Simple Linear Regression)
 - 중다선형 회귀분석 (Multiple Linear Regression)
- ▶ 독립변수가 종속변수에 미치는 영향관계를 설명하거나, 회귀식을 통해 독립변수가 종속변수에 미치는 영향을 예측하는 것
- ▶ 회귀계수 (Regression coefficient)는 이 예언의 값(계수)
- ▶ 상관관계와 비슷하지만 회귀분석은
 - 두 변수간(또는 둘 이상)의 관계에 대한 $y = a + bx$ 라는 직선의 방정식을 구함으로 훨씬 더 정확한 예측 가능
 - b 는 회귀계수, a 는 상수(constant)

예측정확도 평가

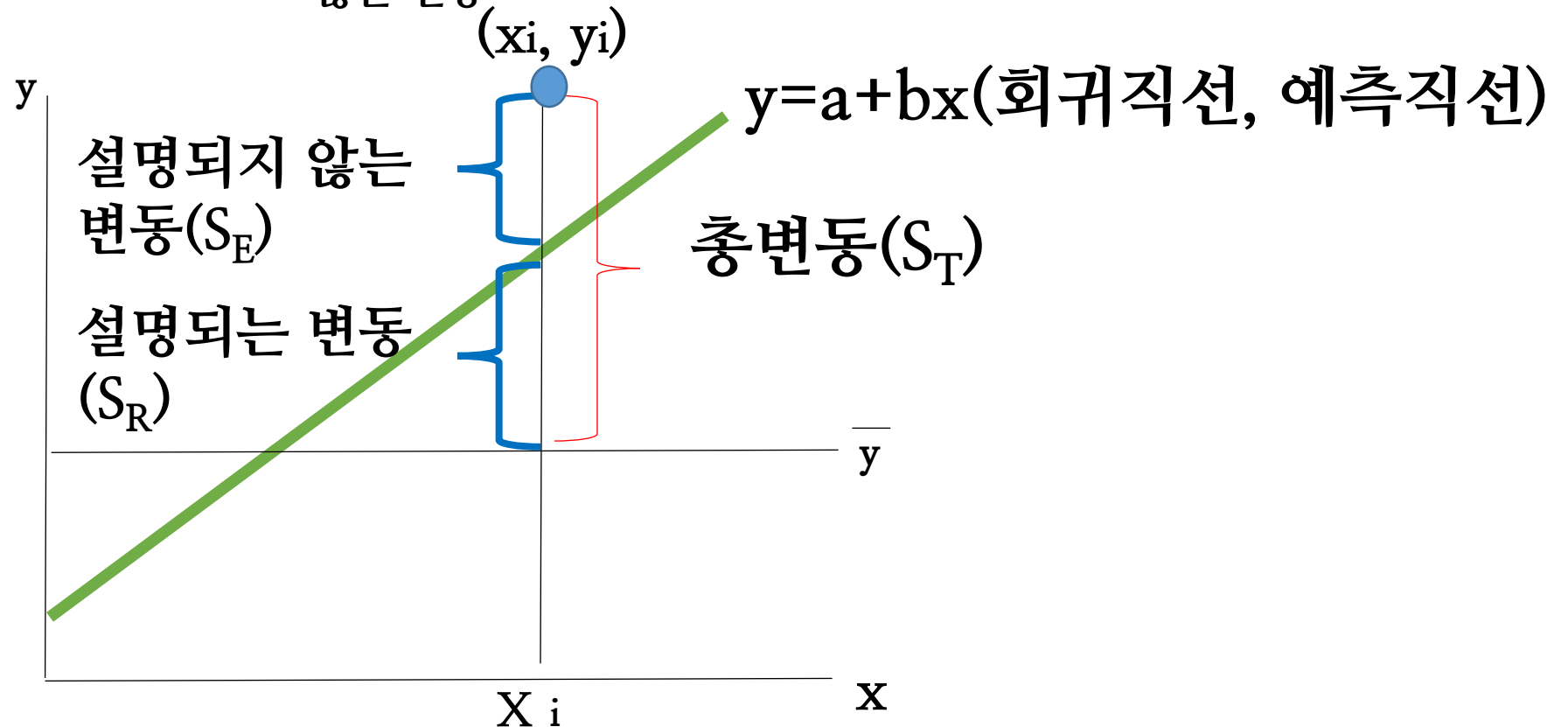
$$S_T = S_E + S_R$$

총변동

회귀에 의해 설명되지
않는 변동

회귀에 의해 설명되는 변동

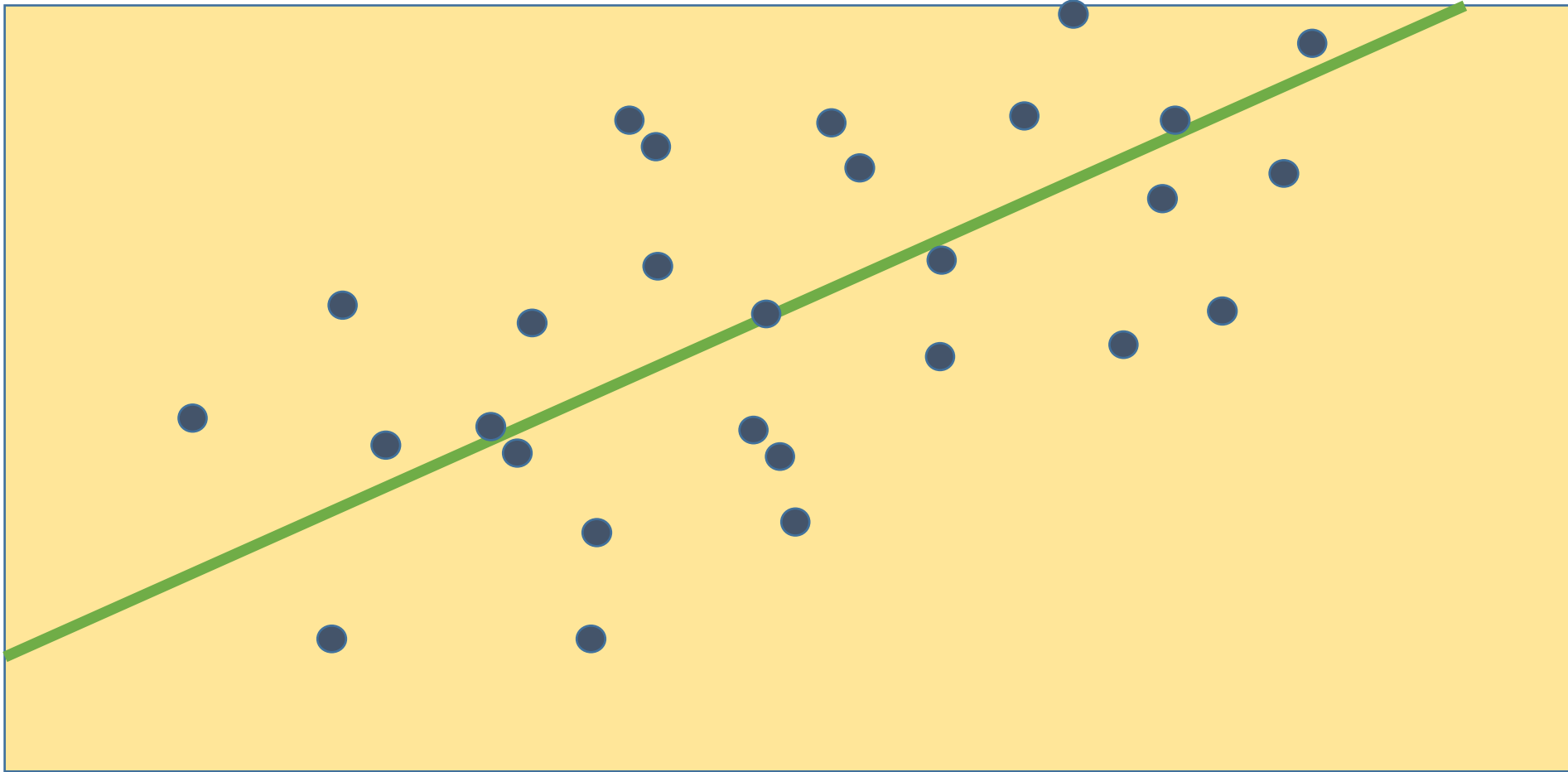
$$R^2 = \frac{S_R}{S_T}$$



예) 학생의 키를 보고 몸무게를 예측할 수 있을까?

y
(몸무게)

종속변수
반응변수
결과변수

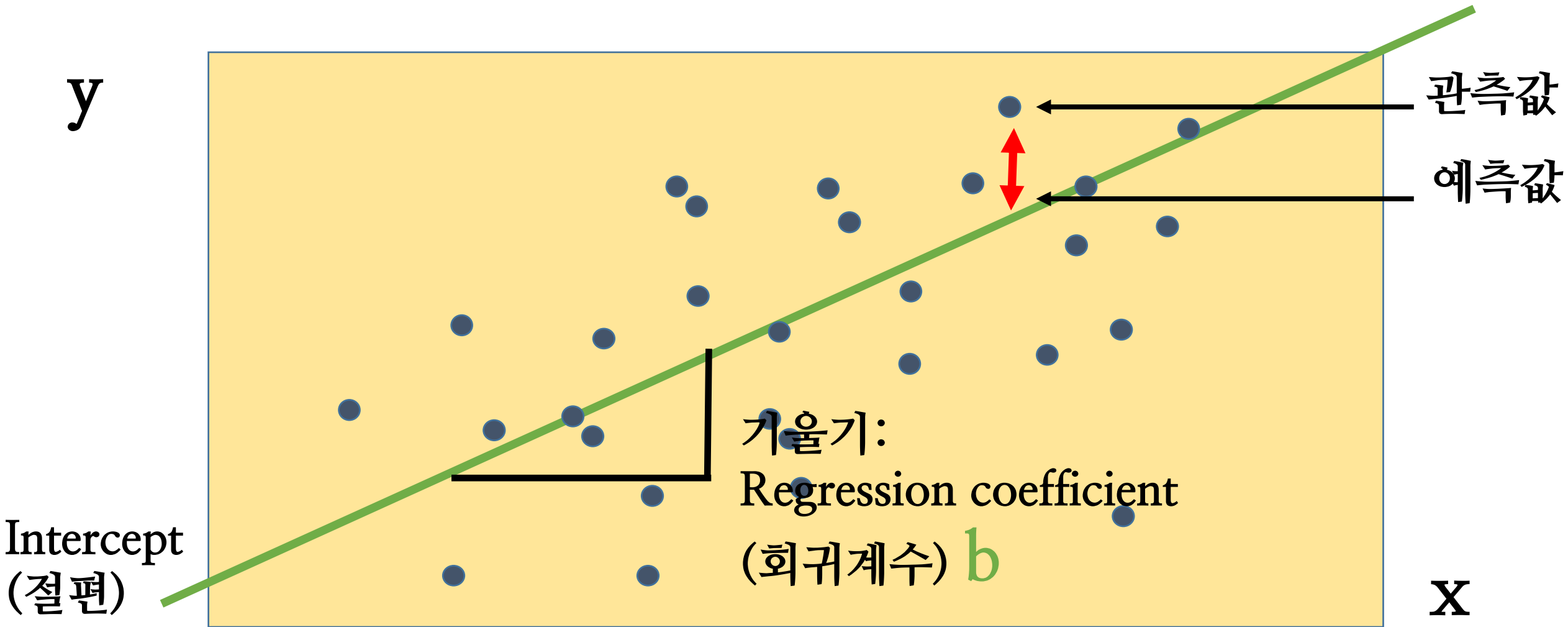


$y=a+bx$ (회귀직선, 예측직선)

X(키)

독립변수
설명변수
예측변수

가장 적합한 직선은 어떻게 얻을 수 있을까??



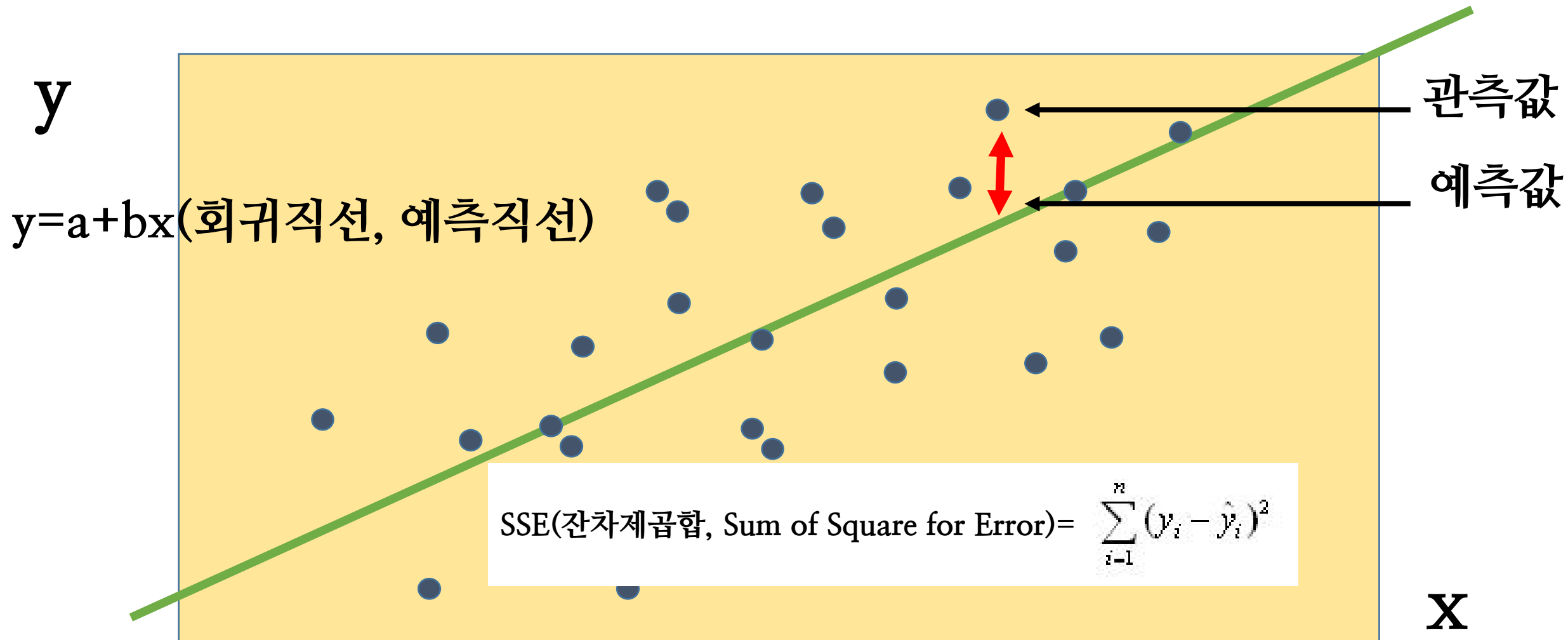
$Y = a + bx_1$

$Y = a_0 + bx_1 + \epsilon$

Prediction error : Residual(잔차)

개별 관측값과 예측값의 차이

가장 적합한 직선은 어떻게 얻을 수 있을까??



최소제곱법(Least square method)을 통해

↕ 모든 세로선의 길이(잔차)의 제곱의 합(SSE)이 최소가 되도록 a 와 b 값을 구해야 함

다중회귀분석

예) 8가족의 신용카드 이용실태에 대한 연구결과

연구 목적 : 어떤 요인들이 이용하고 있는 신용카드의 수에 영향을 미치고 있는가를 결정하기 위함
세가지 가능한 요인들: 가족규모, 가구소득, 자동차 보유대수

Family ID	Number of Credit Cards Used (Y)	Family Size (V ₁)	Income (\$000) (V ₂)	Automobiles Owned (V ₃)
1	4	2	14	1
2	6	2	16	2
3	6	4	14	2
4	7	4	17	1
5	8	5	18	3
6	7	5	21	2
7	8	6	17	1
8	10	6	25	2

종속변수

독립변수

(Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis 7th edition: Pearson 2010. 158 p.)

Multiple regression

- Multiple regression equation

$$= b_0 + b_1V_1 + b_2V_2 + e$$

↑
Prediction error : Residual

Least square method(최소제곱법)를 통해
sum of squared errors
(SSE : 오차의 제곱의 합)을 최소화하는 회귀식 설정

회귀가정의 검증

Linear regression 기본 가정

OVER VIEW

1. 분석전 점검

- 자료에 대한 가정 : 이상점 점검, 정규분포
- Multicollinearity 점검 → 상관분석으로 확인

2. Residual assumption에 대한 점검 후 모형 세우기

Residual analysis

- 1) Normality(정규성) 점검 : residual의 분포는 평균=0인 정규분포
- 2) Independence(독립성) 점검 : data point는 독립적
- 3) Homoscedasticity(등분산) 점검 : 분산이 같아야 함

분석 전 점검

자료에 대한 가정

이상값

정규분포

선형성

다중공선성

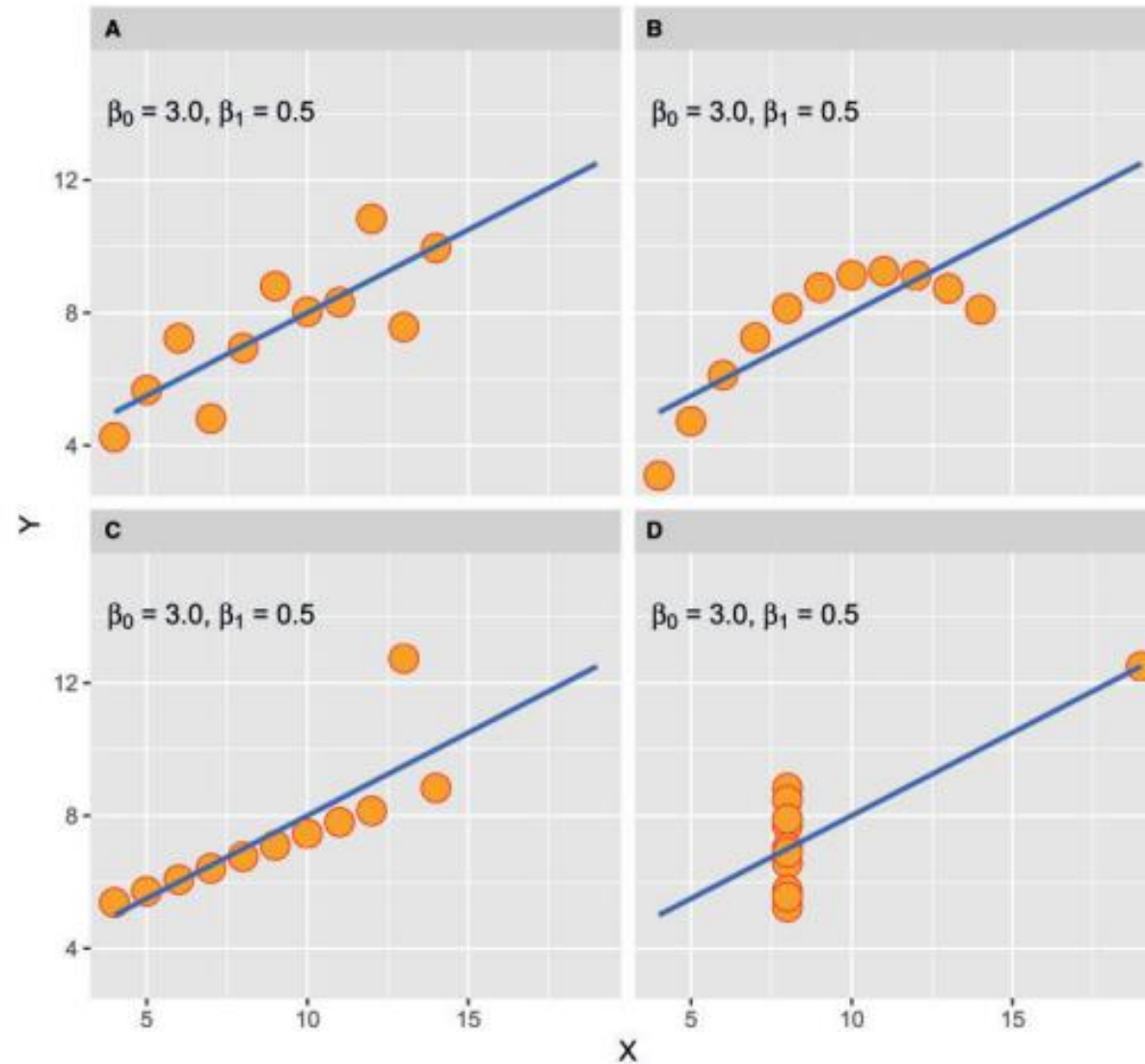


Figure 1: Scatterplots of 4 different datasets known as Anscombe's quartet [8]. Each dataset consists of 11 data points (orange points) and has nearly identical statistical properties, including means, sample variances, the Pearson's sample correlation statistic and linear regression line (blue lines; $\beta_0 = 3, \beta_1 = 0.5$). This well-known quartet highlights the importance of graphing data prior to analysis. (A) A simple linear regression model between the outcome Y and a single continuous predictor variable X. (B) The inappropriateness of ignoring non-linear terms in the regression model. (C and D) An outlier point and leverage point, respectively.

Hickey GL, Kontopantelis E, Takkenberg JJM, Beyersdorf F. Statistical primer: checking model assumptions with regression diagnostics. *Interact CardioVasc Thorac Surg* 2019;28:1–8.

다중공선성 (Multicollinearity)

- Collinearity : 두 독립변수 간의 상관관계로 측정된 연관성
- Multicollinearity : 세 개 이상 간의 상관관계로 측정된 연관성
- 다중공선성이 미치는 영향
 - Predictive power 감소 (하나의 독립변수가 다른 독립변수들과 연관되므로)
 - To maximize the prediction :
 - 독립변수들이 다른 독립변수들과 다중공선성이 낮은 관계
 - 종속변수와 상관관계 높은 독립변수를 찾아내야 함.

◎ 분석 전: 가장 단순하고 분명한 수단은 독립변수들의 상관관계행렬 검토

◎ 회귀 분석 결과: 변수의 공차(tolerance), 분산팽창계수(variance inflation factor, VIF) 검토

(공차: 독립변수에 의해 설명되지 않은 종속변수의 변이 비율)

잔차 분석을 통한 가정진단

회귀모형 기본 가정

잔차(오차항)

정규성

잔차(오차항)

독립성

잔차(오차항)

등분산성

1. 정규성

잔차(오차항)

2. 독립성

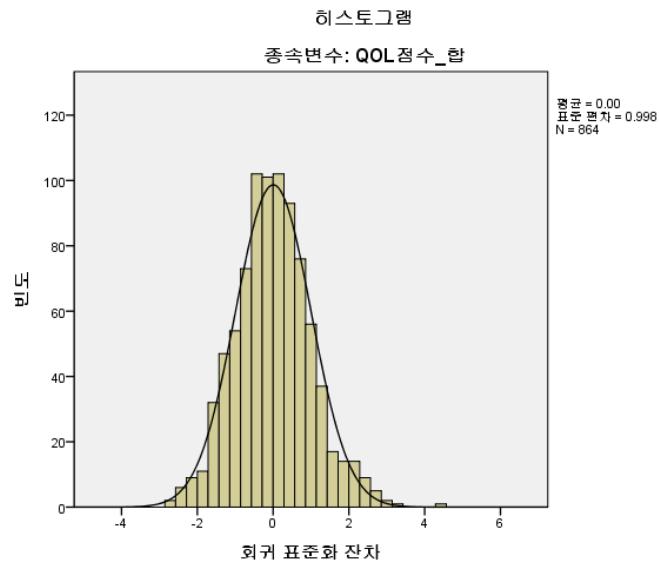
잔차(오차항)

3. 독립성

잔차(오차항)

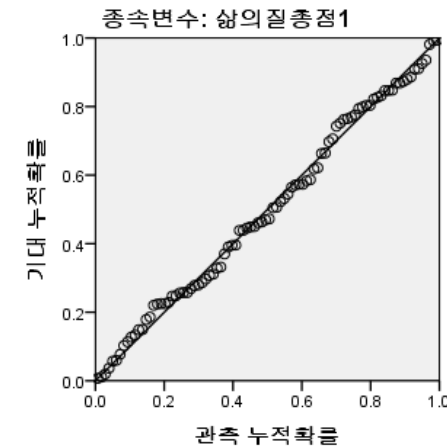
회귀모형 기본 가정

- 두 변수 관계가 선형이고 종속변수와 독립변수의 개개 값에 정규분포하고 있다면, 잔차의 분포는 정규곡선에 근접해야 함
- 표준화 잔차의 분포는 히스토그램, 정규 P-P곡선을 통해 확인



표준화 잔차의 분포위에 정규분포 곡선을 그렸을 때 그 형태가 비슷

회귀 표준화 잔차의 정규 P-P 도표



관측된 확률에 거의 근접하여 증가

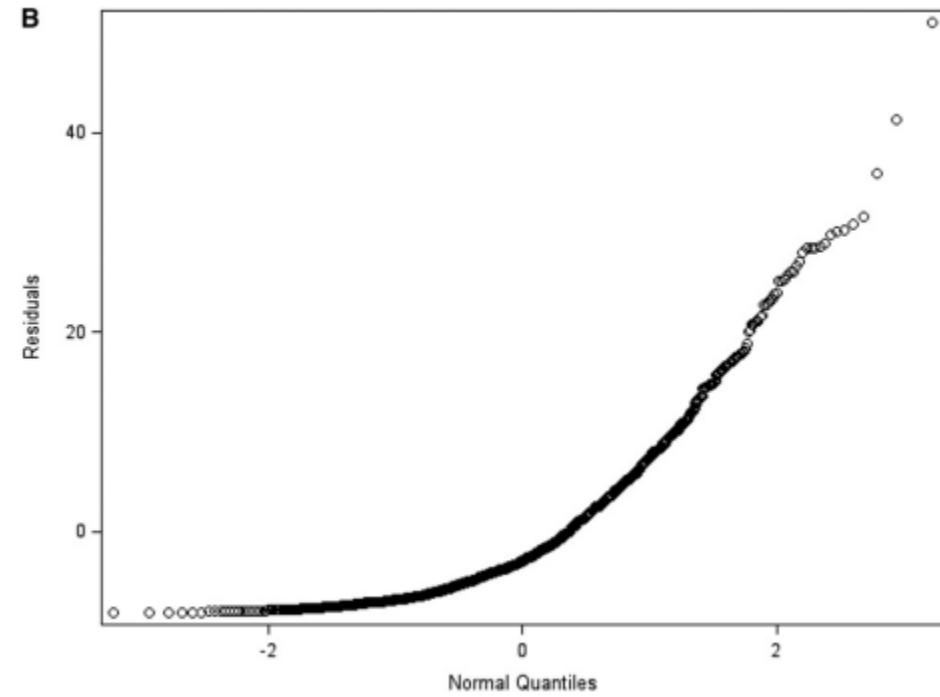
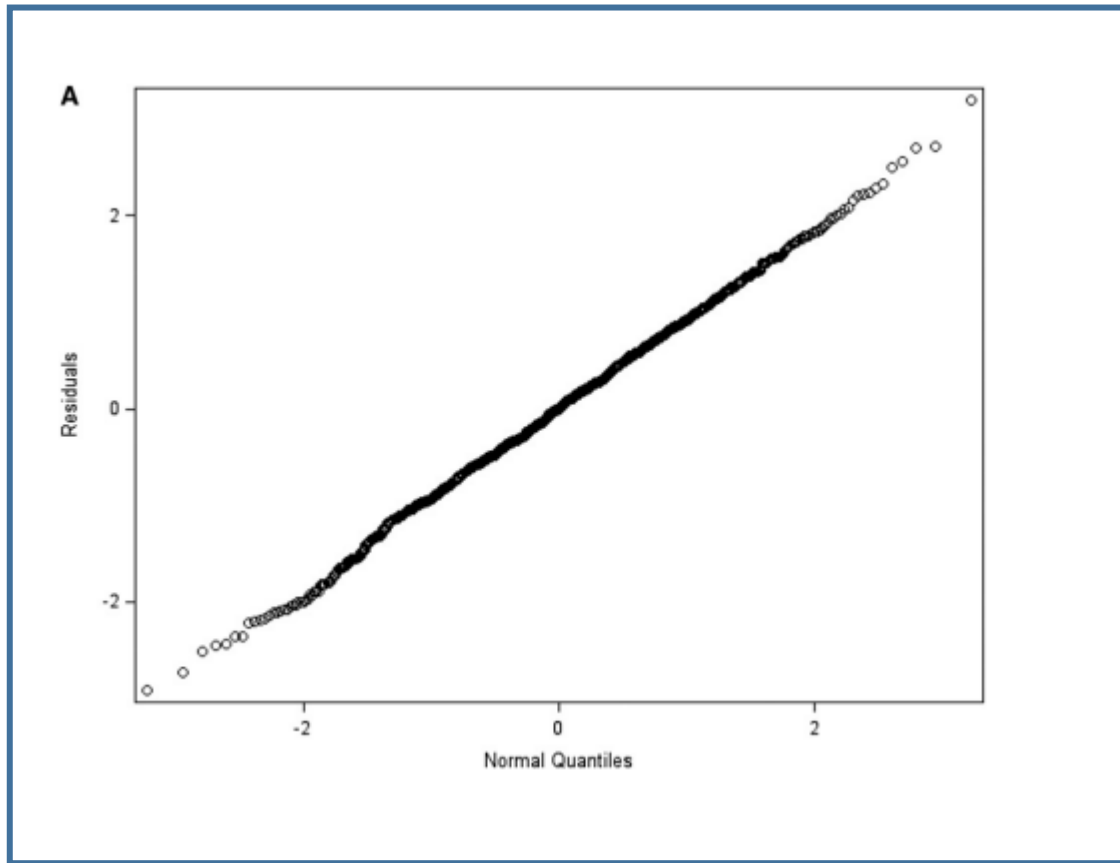


FIGURE 2 Residuals normally distributed, quantile-quantile plot (A). Residuals not normally distributed, quantile-quantile plot (B).

Barker, L. E., & Shaw, K. M. (2015). Best (but oft-forgotten) practices: checking assumptions concerning regression residuals. *American Journal of Clinical Nutrition*, 102(3), 533–539. doi:10.3945/ajcn.115.113498

1. 정규성

잔차(오차항)

2. 독립성

잔차(오차항)

3. 등분산성

잔차(오차항)

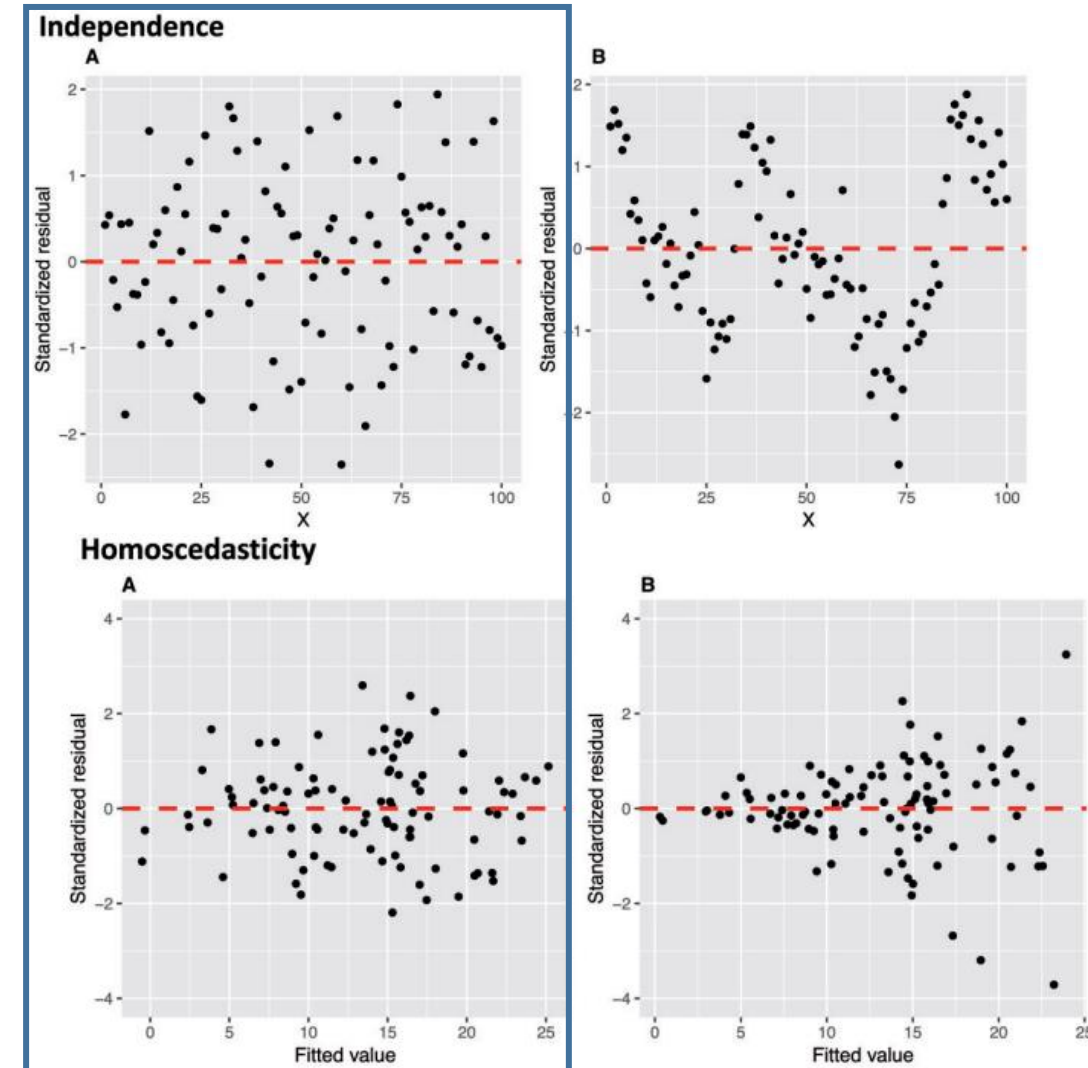
독립성

- 잔차들은 서로 영향을 받지 않고 독립적일 것
- 산점도를 통해 확인

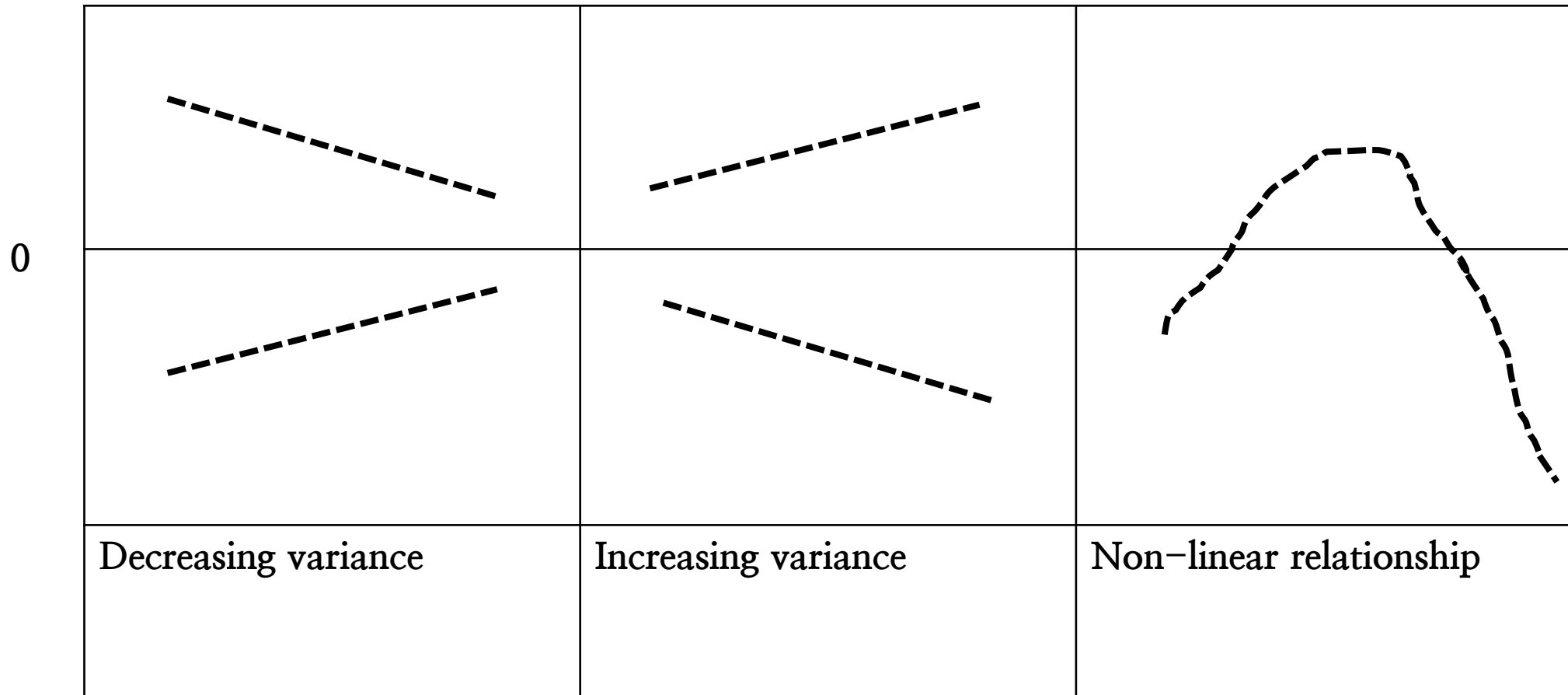
등분산성

- 모든 독립변수의 값에서 잔차의 분산은 같음
- 산점도를 통해 확인

Hickey GL, Kontopantelis E, Takkenberg JJM, Beyersdorf F . Statistical primer: checking model assumptions with regression diagnostics. Interact CardioVasc Thorac Surg 2019;28:1-8.



- 다음과 같은 경향을 보이면 등분산이 같다는 가정에 위반



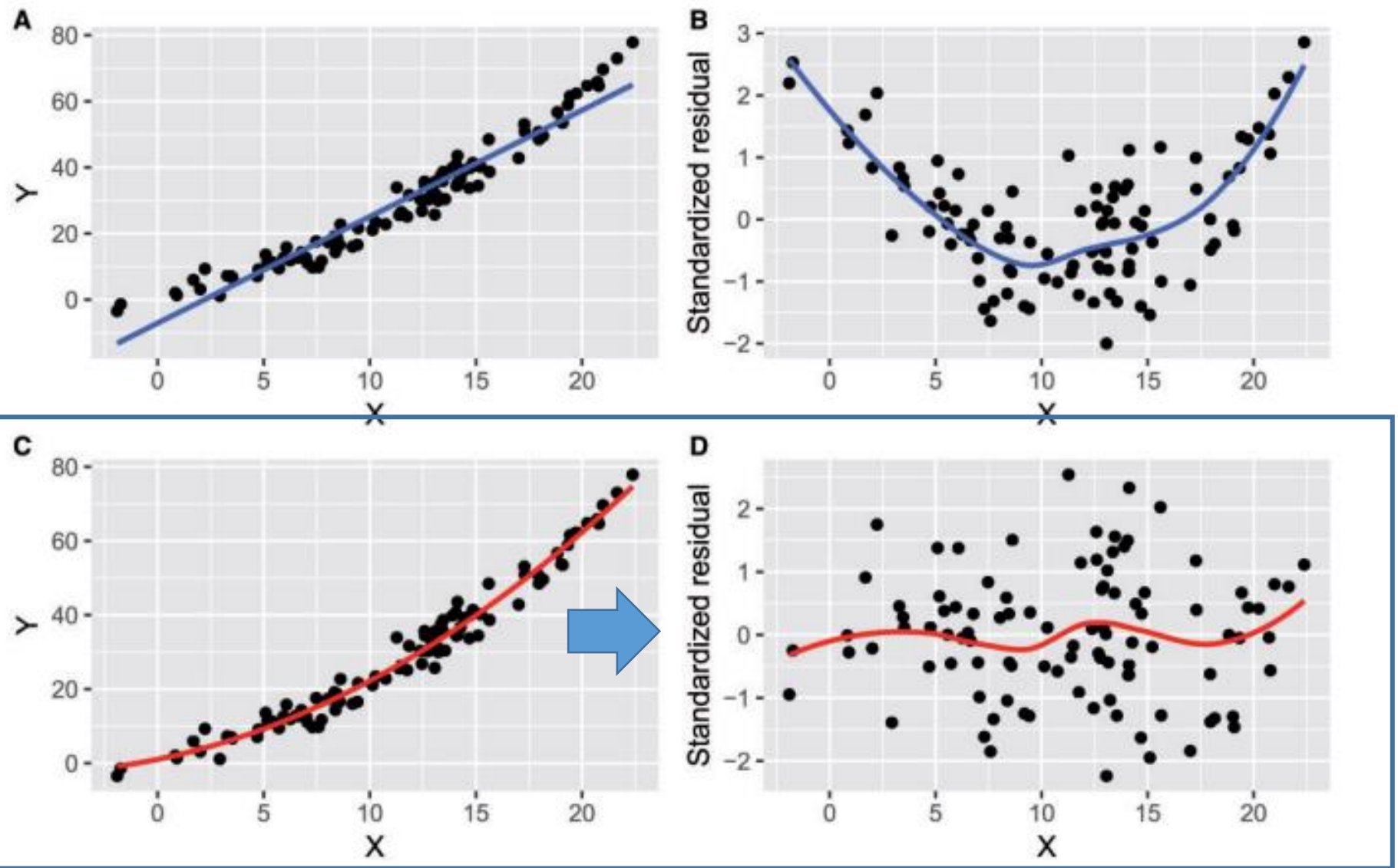


Figure 2: (A) A scatterplot of data with a linear regression line fitted of the form $Y = \beta_0 + \beta_1 X + \epsilon$. (B) A scatterplot of the predictor variable (X) versus the residuals from the model in (A). A clear U-shape pattern is displayed indicating a possible quadratic relationship. (C) As per (A) with a linear regression line fitted of the form $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$. (D) As per (B) but with the residuals from the model with a quadratic term. No obvious trend in the residuals remains.

Hickey, G. L., Kontopantelis, E., Takkenberg, J. J. M., & Beyersdorf, F. (2019). Statistical primer: checking model assumptions with regression diagnostics. *Interactive Cardiovascular and Thoracic Surgery*, 28(1), 1–8.
doi:10.1093/icvts/ivy207

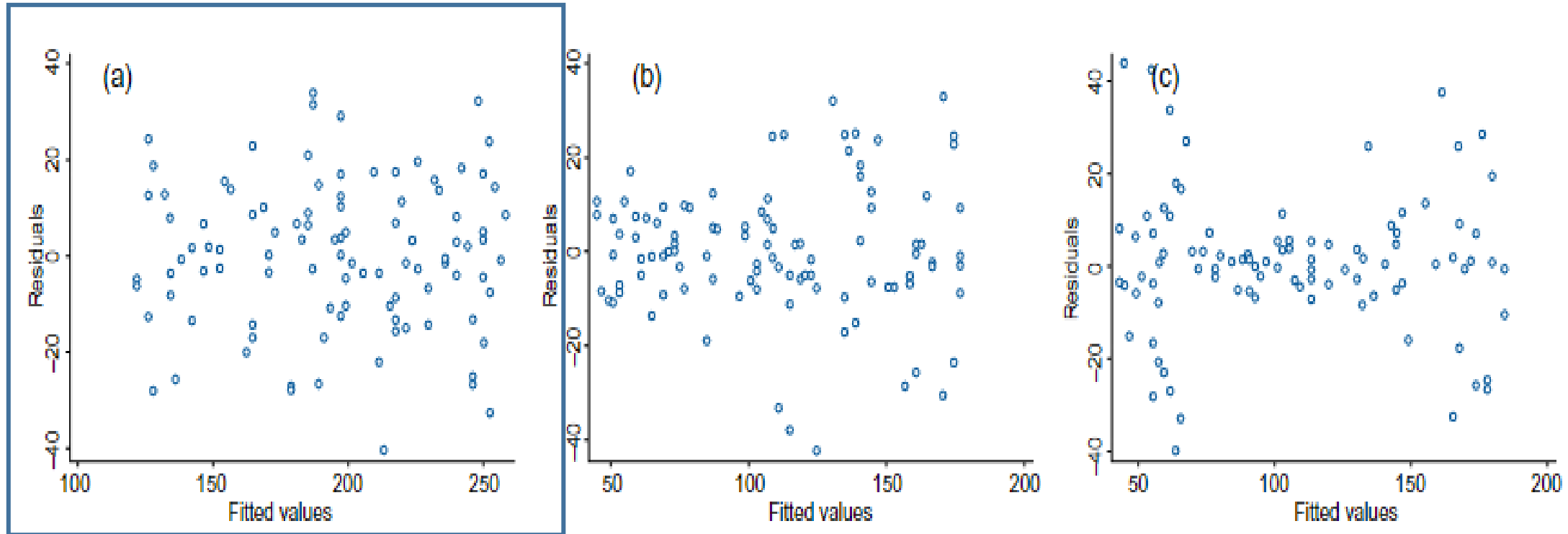


Figure 1. Scatter plots and histograms of data and residuals. (a–c) Residuals versus fitted (predicted values). (a) ‘Well-behaved’ residual pattern, demonstrating randomness. (b and c) Non-constant variance (heteroscedasticity), violating the assumption of constant variance.

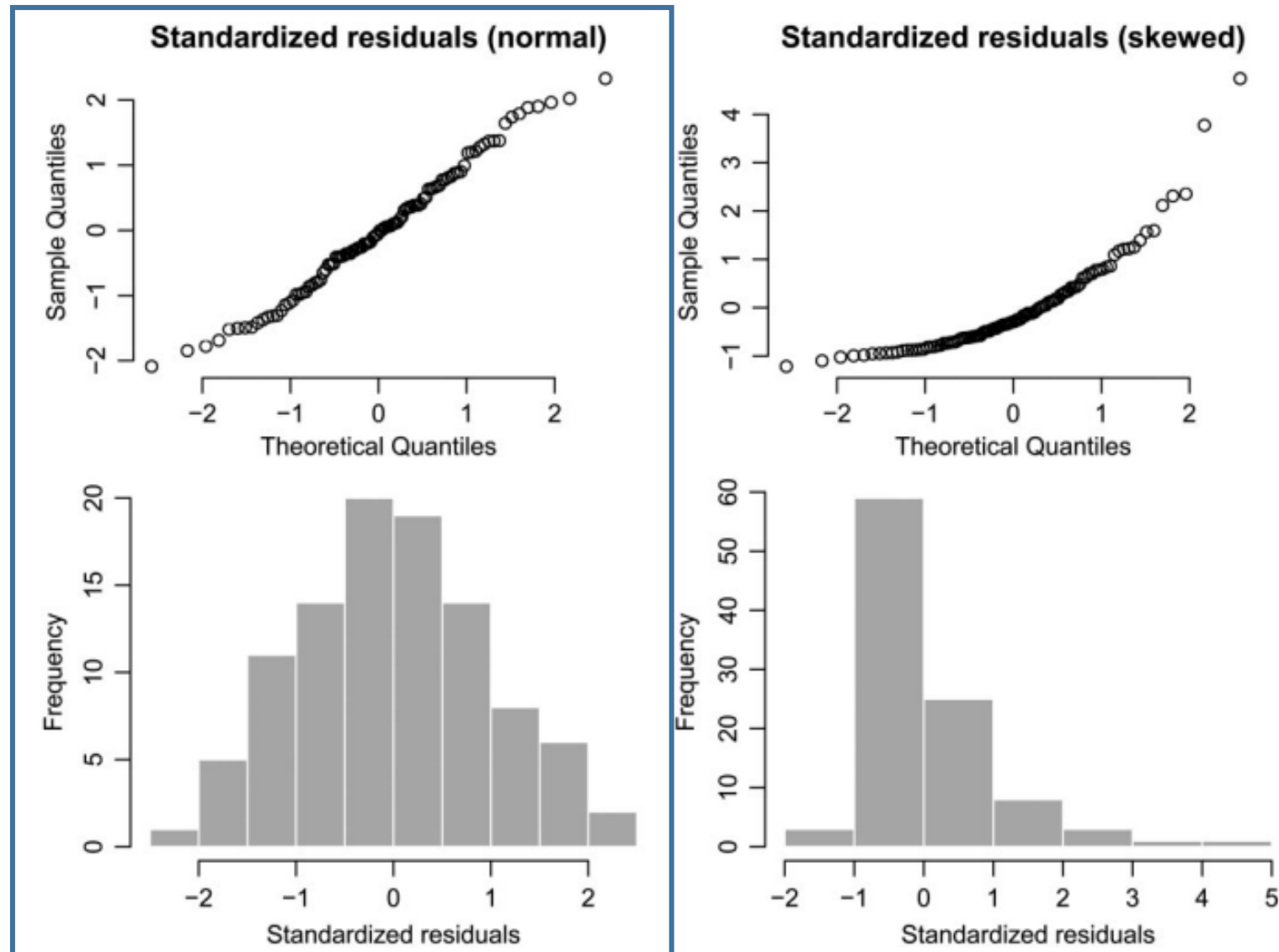
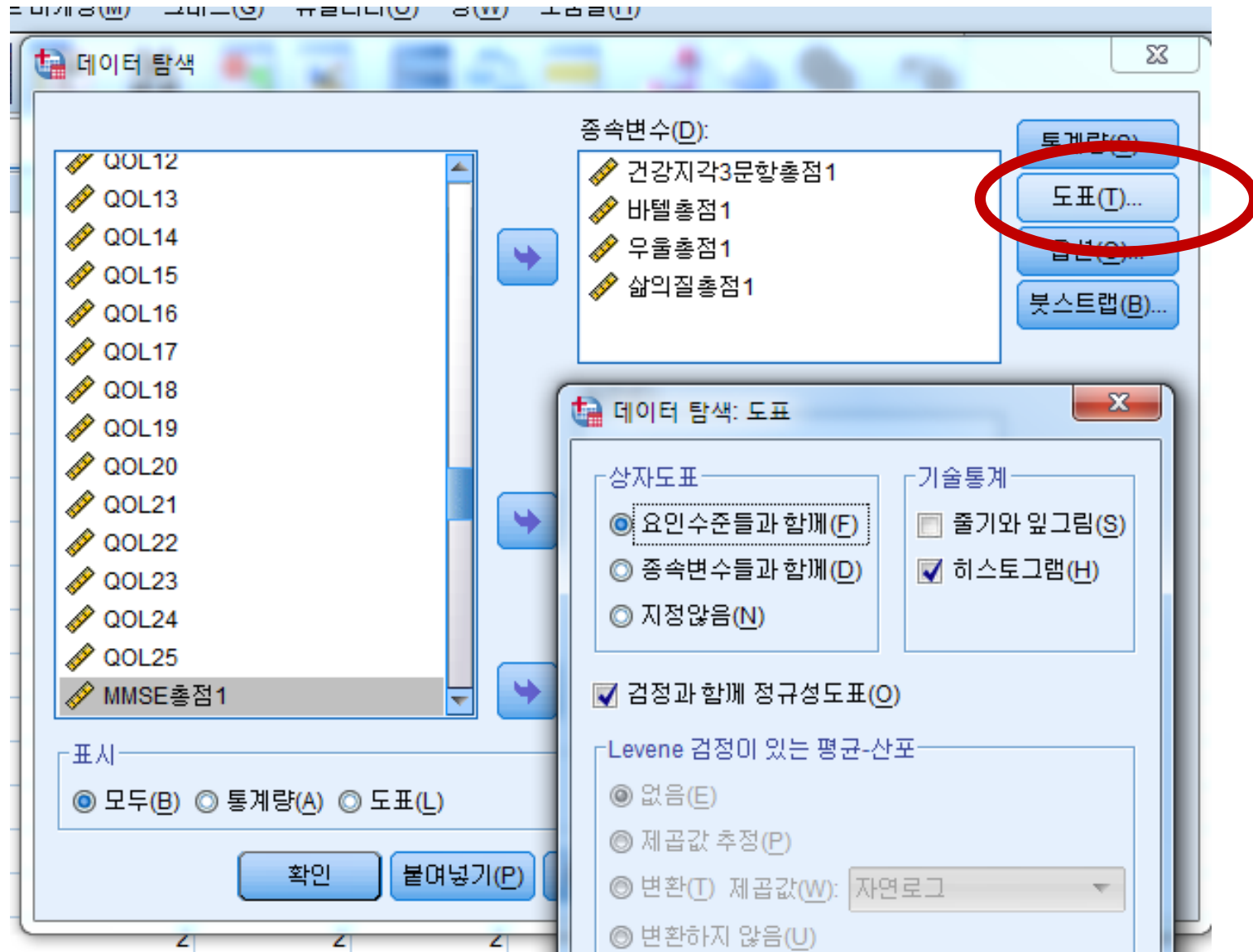


Figure 4: Q-Q plots (top row) and histograms (bottom row) of residuals from a model with approximately normally distributed residuals (left column) and skew-distributed residuals (right column).

Hickey, G. L., Kontopantelis, E., Takkenberg, J. J. M., & Beyersdorf, F. (2019). Statistical primer: checking model assumptions with regression diagnostics. *Interactive Cardiovascular and Thoracic Surgery*, 28(1), 1–8.
doi:10.1093/icvts/ivy207

SPSS 실습

데이터탐색(연속변수): 분석-기술통계량-데이터탐색



정규성 검정

- 통계적 검정

- 정규분포와의 차이에 대한 유의수준 계산
 - Shapiro-Wilks / Kolmogorov-Smirnov
 - 정규성 검정의 가설 검정

귀무가설(H_0): 자료는 정규분포를 따른다.
대립가설(H_1): 자료는 정규분포를 따르지 않는다.

- 탐색적 검정

- 그래프에 의한 정규분포
 - 막대그림표, 정규확률도면(normal probability plot), 왜도(skewness), 첨도(kurtosis)

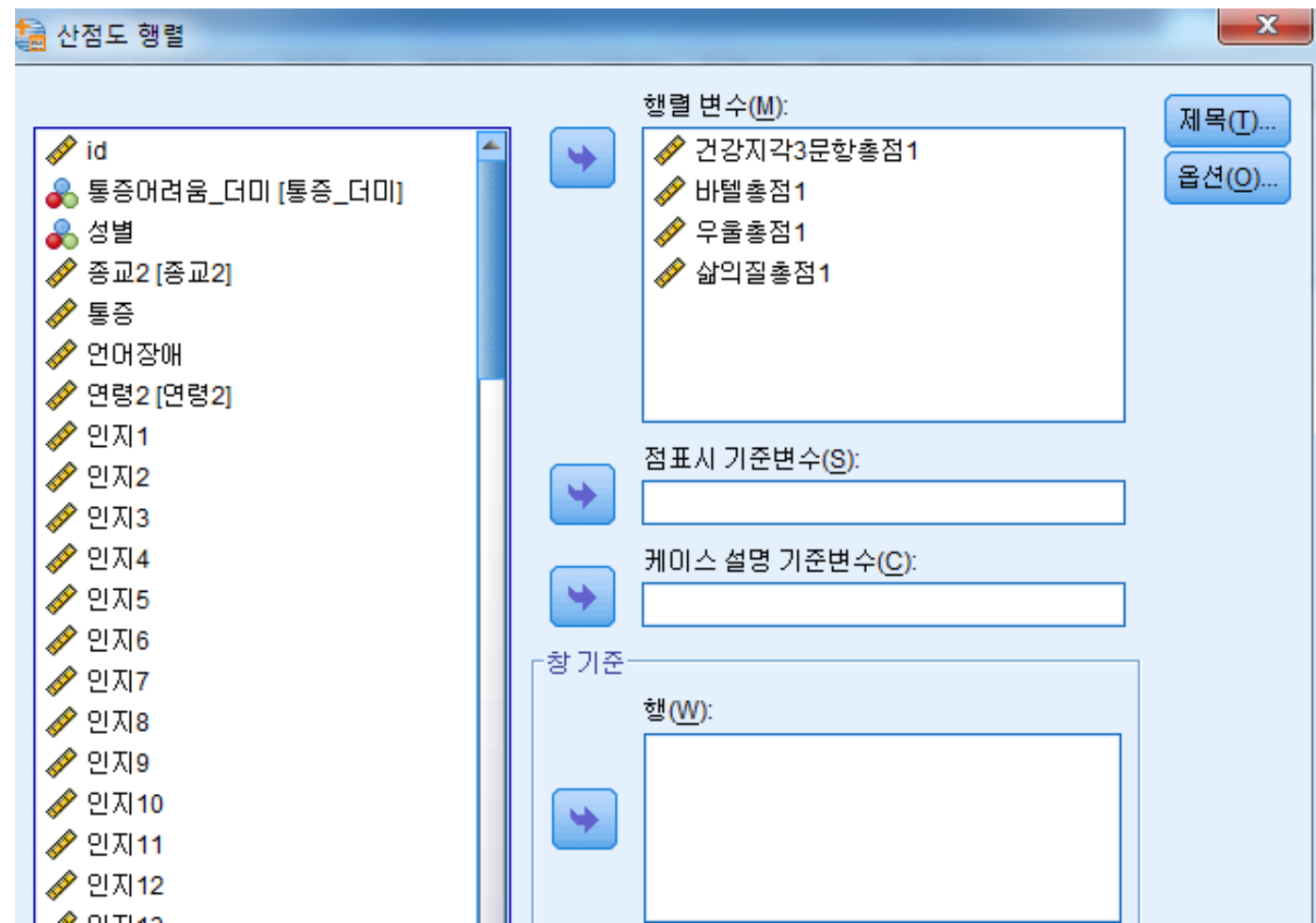
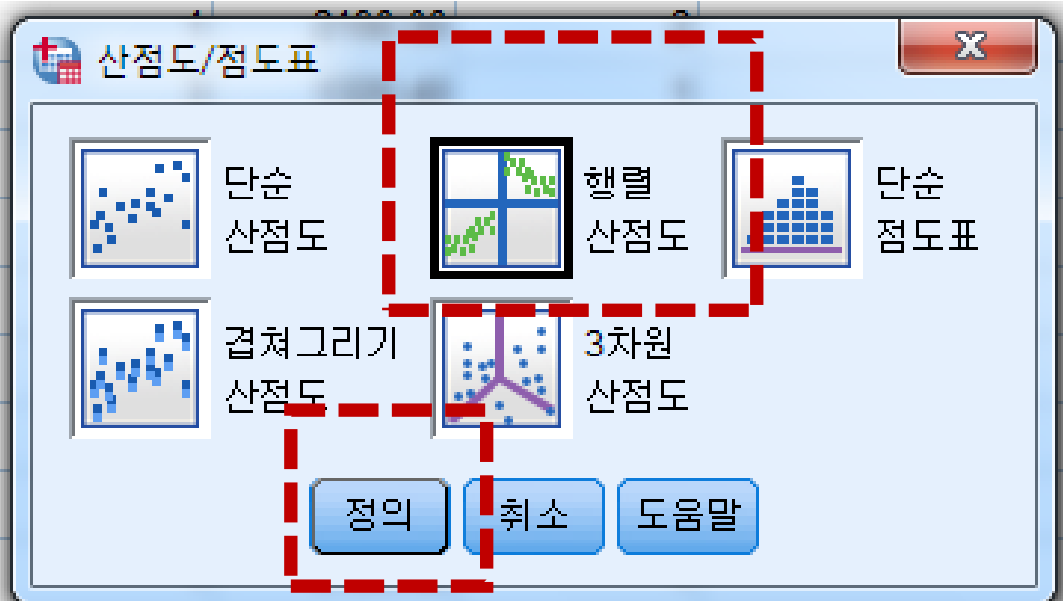
행렬산점도 : 그래프-래거시 대화 상자-산점도

□ 선형성, 이상값, 상관성, 다중공선성확인

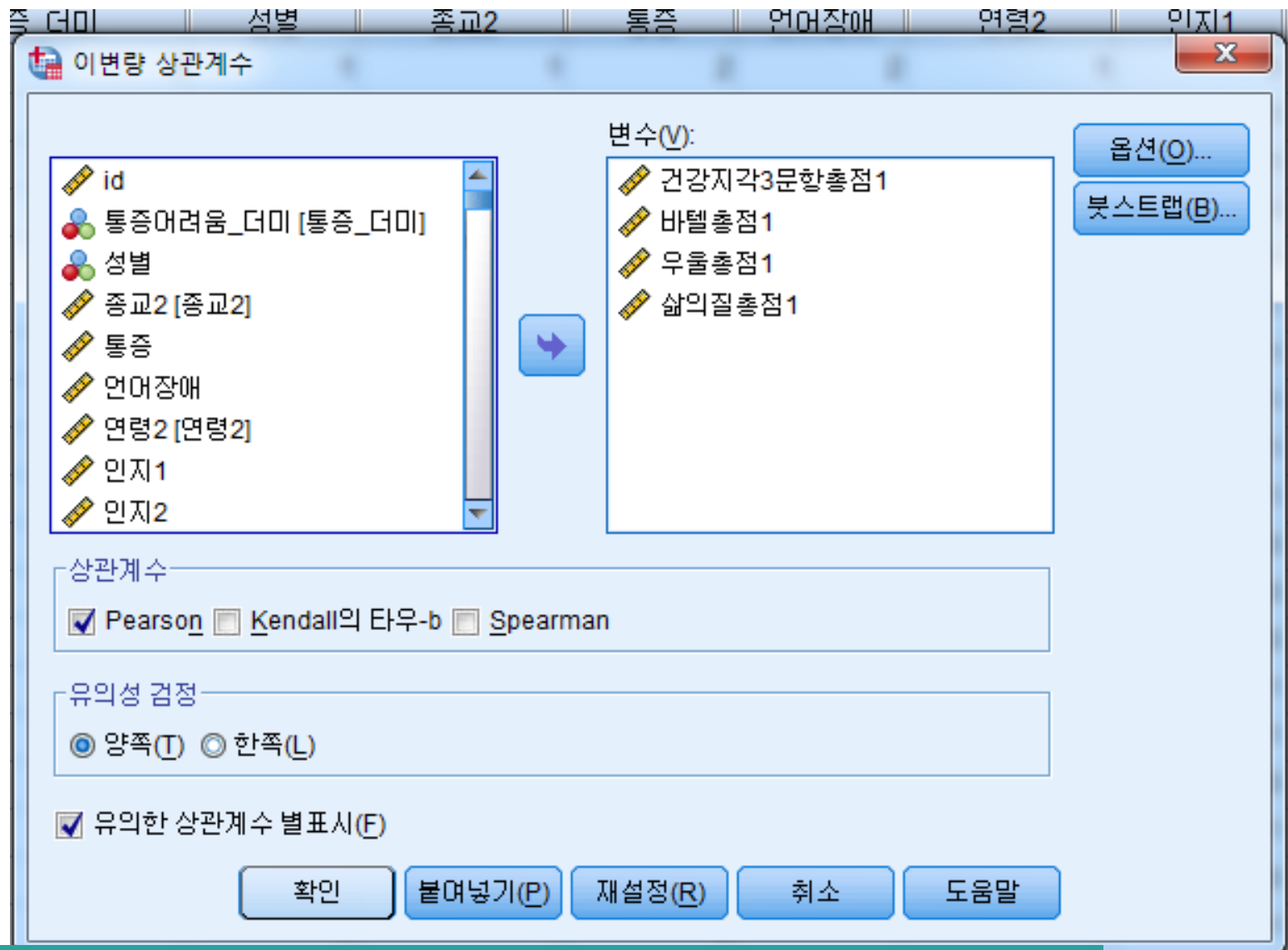
- 산점도는 변수와 표본의 개수가 비교적 적은 경우에 유용
- 변수와 표본의 수가 많으면 상관분석이 효율적

행렬산점도 그리기

- 그래프-레거시 대화상자-산점도/점도표-**행렬산점도**→정의



다중공선성 : 상관분석(분석-상관분석-이변량상관계수)



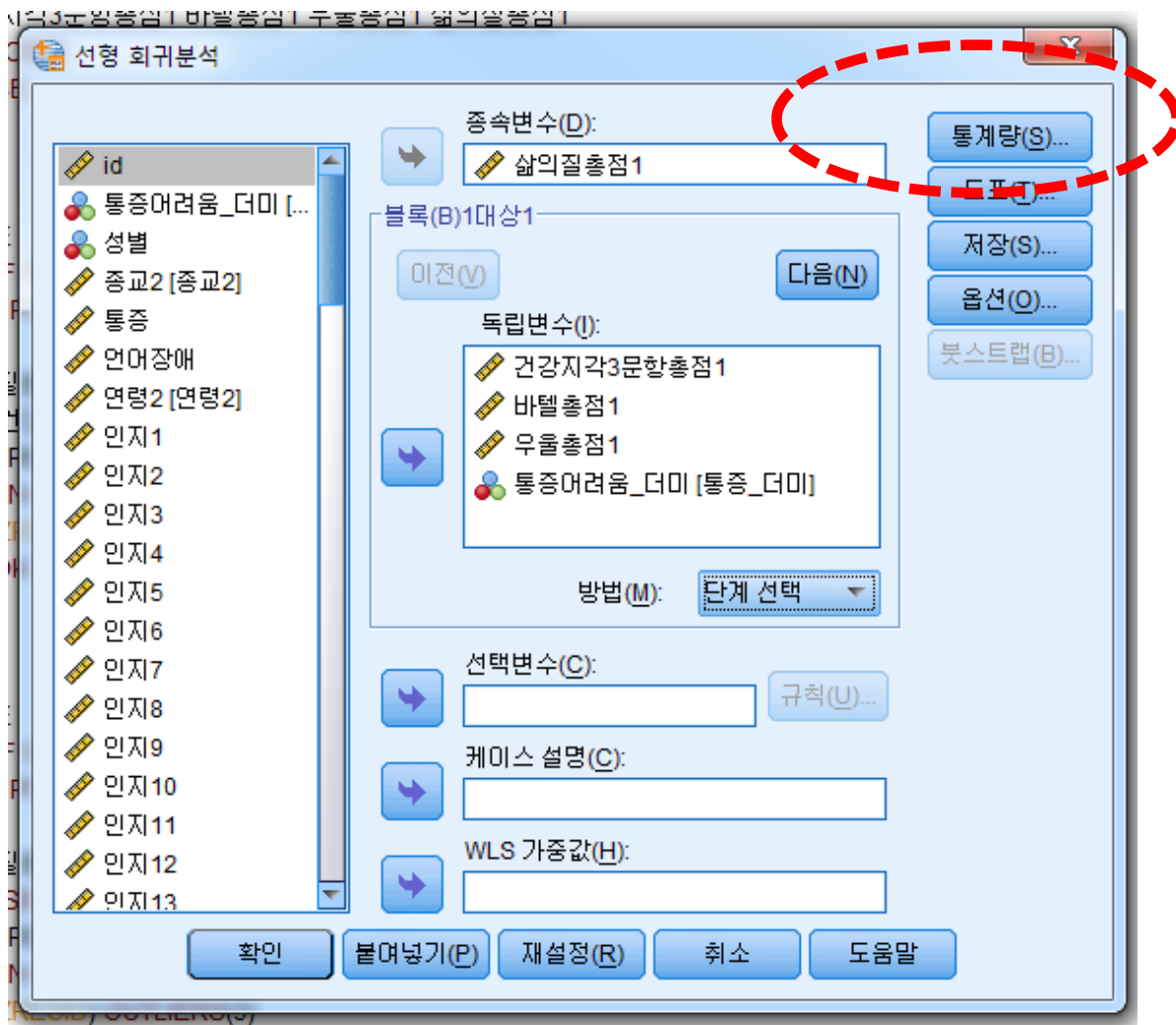
상관계수

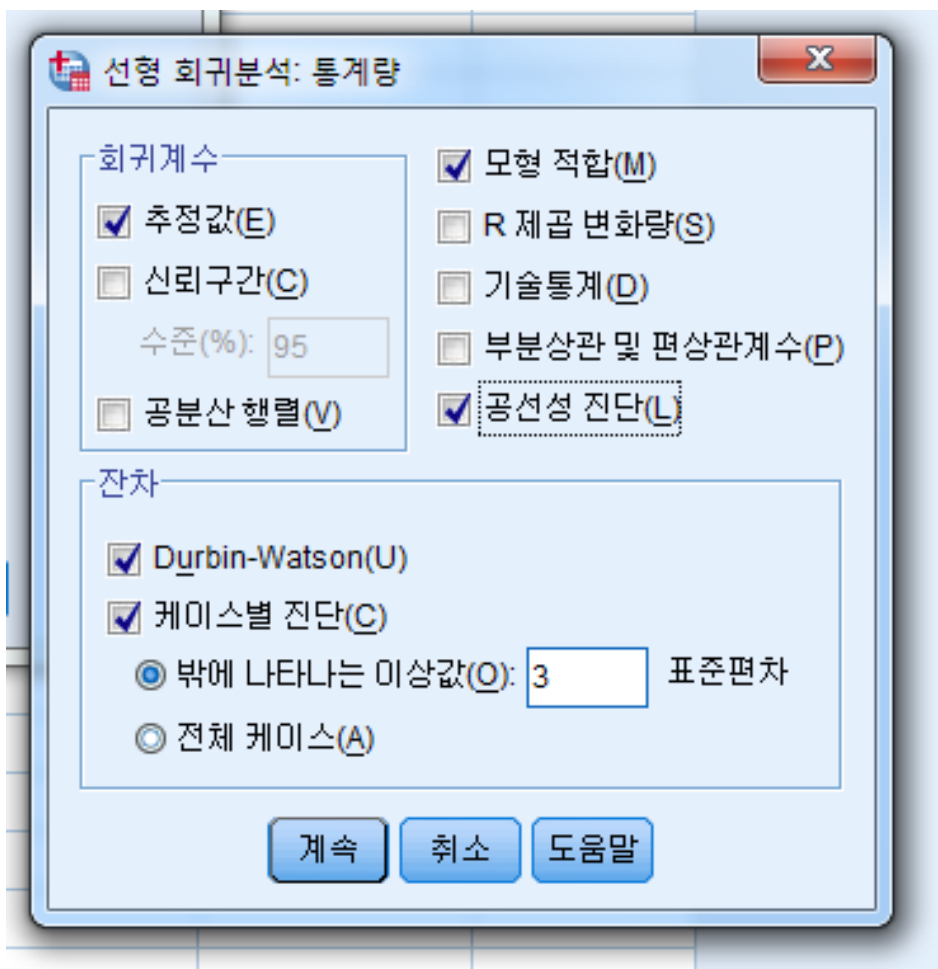
		건강지각3문항 총점1	바텔총점1	우울총점1	삶의질총점1
건강지각3문항총점1	Pearson 상관계수	1	-.131	.460**	-.563**
	유의확률 (양쪽)		.214	.000	.000
	N	92	92	92	92
바텔총점1	Pearson 상관계수	-.131	1	-.372**	.343**
	유의확률 (양쪽)	.214		.000	.001
	N	92	92	92	92
우울총점1	Pearson 상관계수	.460**	-.372**	1	-.724**
	유의확률 (양쪽)	.000	.000		.000
	N	92	92	92	92
삶의질총점1	Pearson 상관계수	-.563**	.343**	-.724**	1
	유의확률 (양쪽)	.000	.001	.000	
	N	92	92	92	92

** . 상관계수는 0.01 수준(양쪽)에서 유의합니다.

- 종속변수(QOL)와 독립변수들간의 상관관계: 모든 변수가 유의한 상관관계
회귀분석을 실시하면 이들 독립변수가 종속변수에 유의한 영향을 줄 것으로 예상
- 독립변수들 간의 상관관계: 모든 | r | 값이 .8 보다 작으므로 **다중 공선성 문제**에서 비교적 여유로움

다중회귀분석 : 분석-회귀분석





통계량

- 회귀계수의 추정값, 모형적합, 공선성진단 선택
- 잔차

- Durbin-Watson

- Independence 점검

- 0-4 범위
 - 2주위의 값이 나오면 data point가 독립적
 - 0에 가까우면 강한 양의 상관관계
 - 4에 가까우면 강한 음의 상관관계

- Casewise diagnostics

- Outlier의 점검

- 전체 사례에 대해
 - 표준편차 ± 3 벗어나는 경우,
cook통계량 ± 1 벗어나는 경우 극단치로 예상됨

□ 도표

■ Y축

- 표준화된 잔차(Standardized residual:*ZRESID)

■ X축

- 표준화된 예측값(Standardized predicted value:*ZPRED)

→ 잔차의 등분산성 검증

■ 표준화된 잔차도표

- 히스토그램
- 정규확률도표 선택

→ 잔차의 정규성 검증



선형 회귀분석: 저장

예측값

☐ 비표준화(U)

☐ 표준화(A)

☐ 수정된(J)

☐ 평균예측 표준오차(P)

잔차

☐ 비표준화(U)

☒ 표준화(A)

☐ 스튜던트화(S)

☐ 삭제된 잔차(D)

☐ 삭제된 스튜던트화 잔차(E)

거리

☒ Mahalanobis의 거리(H)

☒ Cook의 거리(K)

☒ 레버리지 값(L)

영향력 통계량

☐ DFBETA(B)

☐ 표준화 DFBETA(Z)

☐ DFFIT(F)

☒ 표준화 DFFIT(T)

☐ 공분산 비율(V)

예측 구간

☐ 평균(M) ☐ 개별값(I)

신뢰구간(C): 95 %

계수 통계량

☐ 상관계수 통계량 만들기(O)

☒ 새 데이터 파일 만들기

데이터 파일 이름(O):

☒ 새 데이터 파일 쓰기

파일(I)...

XML 파일에 모형정보 내보내기

찾아보기(W)...

☒ 공분산행렬 포함(I)

계속 취소 도움말

• 저장

• 잔차

- 표준화 잔차의 값
- 표준화 영향력통계량

• 거리

- Mahalanobis의 거리
- Cook의 거리
- 레버리지 값 선택

→ 각각의 값이 OUTPUT에 제공되지 않고 데이터파일에 저장됨

→ 극단치(Outlier) 탐색

gender	ZRE_1	MAH_1	COO_1	LEV_1	SDF_1	변수
2	.81471	4.96796	.00068	.00584	.06873	
2	-1.40884	4.17998	.00175	.00492	-.11071	
2	.	3.58523	.	.00421	.	
2	-.12707	3.84537	.00001	.00452	-.00964	
2	1.04158	12.67866	.00258	.01492	.13429	
1	-1.03010	4.22905	.00094	.00498	-.08129	
			.00038	.00487	.05155	

Syntax

REGRESSION

/MISSING LISTWISE

/STATISTICS COEFF OUTS R ANOVA COLLIN TOL

/CRITERIA=PIN(.05) POUT(.10)

/NOORIGIN

/DEPENDENT 삶의질총점1

/METHOD=STEPWISE 건강지각3문항총점1 바텔총점1 우울총점1 통증_더미

/SCATTERPLOT=(*ZRESID ,*ZPRED)

/RESIDUALS DURBIN HISTOGRAM(ZRESID) NORMPROB(ZRESID)

/CASEWISE PLOT(ZRESID) OUTLIERS(3)

/SAVE MAHAL COOK LEVER ZRESID SDFIT.

OUTPUT

가정검증 결과만 내용에 포함됨

모형요약

모형 요약^d

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차	Durbin-Watson
1	.724 ^a	.524	.518	7.983	
2	.768 ^b	.591	.581	7.442	
3	.788 ^c	.620	.607	7.206	1.903

a. 예측값: (상수), 우울총점1

b. 예측값: (상수), 우울총점1, 건강지각3문항총점1

c. 예측값: (상수), 우울총점1, 건강지각3문항총점1, 통증어려움_더미

d. 종속변수: 삶의질총점1

Durbin-Watson : Independence의 점검

0-4의 범위이며
2주위의 값이므로 독립의 가정 만족

계수

계수^a

모형	비표준화 계수		표준화 계수	t	유의확률	공선성 통계량	
	B	표준오차	베타			공차	VIF
(상수)	74.106	3.371		21.981	.000		
3 우울총점1	-1.674	.213	-.582	-7.857	.000	.787	1.271
건강지각3문항총점1	-1.097	.297	-.274	-3.689	.000	.782	1.279
통증어려움_더미	-5.130	1.952	-.174	-2.628	.010	.984	1.016

a. 종속변수: 삶의질총점1

〈공선성 판단을 위한 일반적 기준〉

- Tolerance(공차한계) : .10이상
- VIF(variance inflation factor, 분산팽창요인) : 10 이하일때 다중공선성 낮음

*독립변수들끼리의 상관성(공선성)이 없어야 회귀분석결과 신뢰

다중공선성 (Multicollinearity) 기준

- 공차한계($\text{Tolerance} = 1 - R^2$) : 0.1 이상
- 분산팽창요인($\text{VIF} = 1 / \text{Tolerance} = 1 / (1 - R^2)$) : 10 이하
(1에 가까울수록 좋음)

→ 충족 시 다중공선성 낮으므로 회귀분석 가능함
(Tolerance 작을 수록 VIF 클수록 다중공선성 있다고 의심)

잔차 통계량

□ 표준화 잔차

± 3 을 넘지 않아야
Outlier가 없다고 판정

□ Cook의 거리

± 1 을 넘지 않아야 outlier
없다고 판정

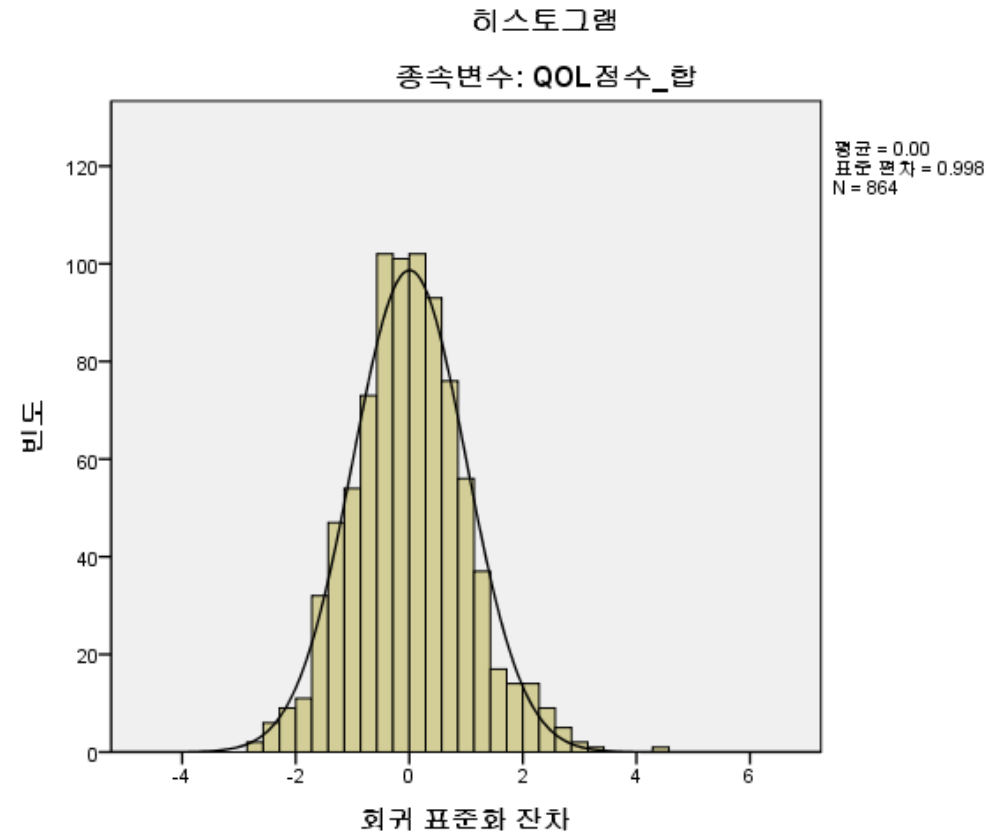
잔차 통계량^a

	최소값	최대값	평균	표준편차	N
예측값	27.21	60.81	41.68	9.076	
표준 오차 예측값	-1.595	2.108	.000	1.000	88
예측값의 표준오차	.897	2.635	1.492	.417	88
수정된 예측값	27.00	60.19	41.71	9.106	88
잔차	-17.400	17.707	.000	7.137	88
표준화 잔차	-2.395	2.438	.000	.983	88
스튜던트 잔차	-2.443	2.481	-.002	1.006	88
스튜던트 잔차	-18.097	18.622	-.031	7.487	88
스튜던트 스튜던트 잔차	-2.520	2.562	-.002	1.019	88
Mahal. 거리	.339	10.459	2.966	2.289	88
Cook의 거리	.000	.104	.012	.019	88
중심화된 레버리지 값	.004	.120	.034	.026	88

a. 종속변수: 삶의질총점1

도표

*ZRESID히스토그램

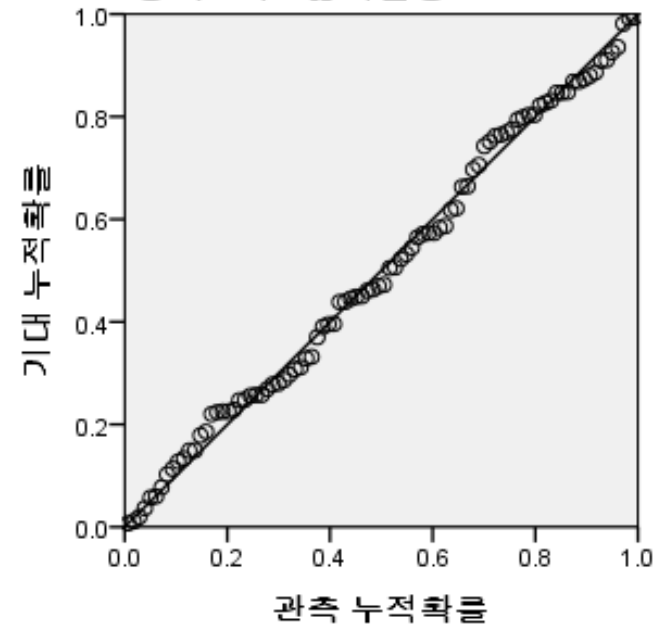


잔차의 Normality의 점검

*ZRESID정규 P-P도표

회귀 표준화 잔차의 정규 P-P 도표

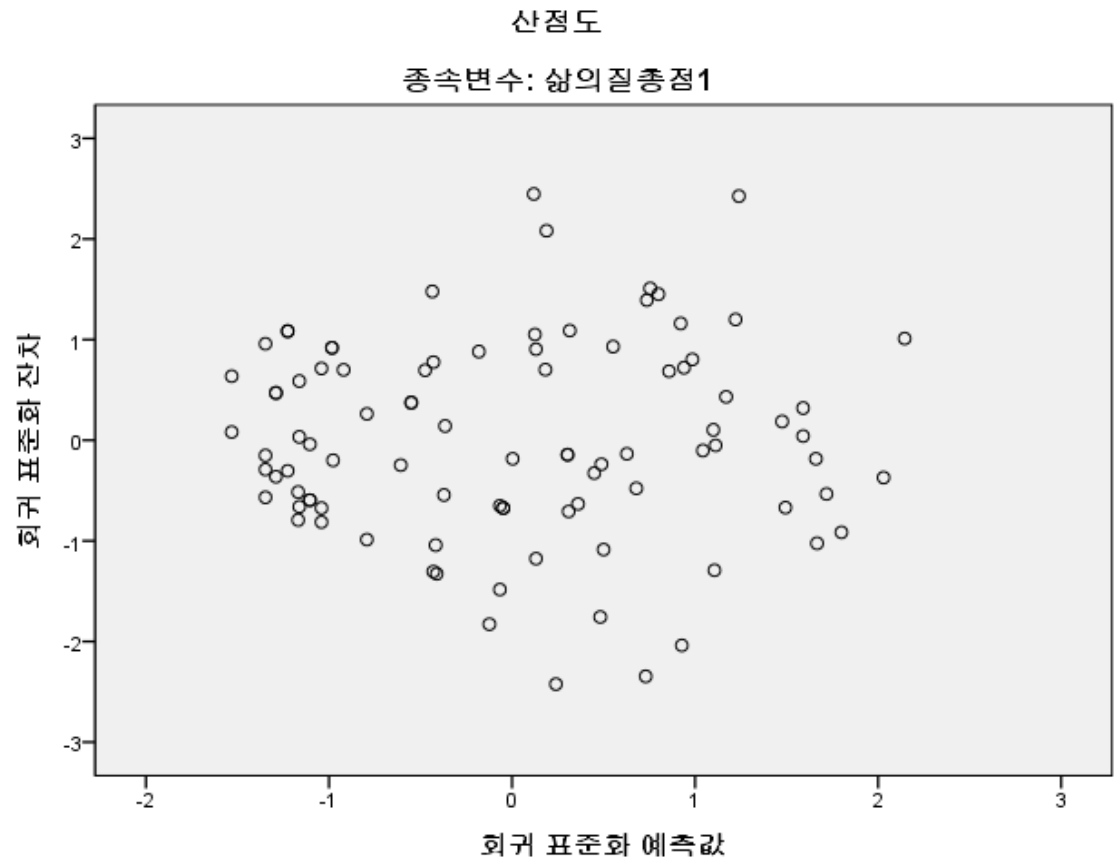
종속변수: 삶의 질 총점1



잔차의 Normality의 점검

관측된 확률에 거의 근접하여 증가하고 있으므로 정규성 보여줌

*ZPRED에 의한 *ZRESID산점도



등분산 가정 점검

수평축과 수직축의 관계가 무작위로 흩어져 있으면 등분산 가정만족

논문의 예



eISSN 2093-758X
J Korean Acad Nurs Vol.51 No.6, 678
<https://doi.org/10.4040/jkan.21131>

감염 전담 병동 간호사의 COVID-19 환자 간호업무수행에 영향을 미치는 요인

김윤선[○] · 김미애[○]
동남권원자력의학원 간호부

5. 대상자의 COVID-19 간호업무수행에 영향을 미치는 요인

대상자의 COVID-19 환자 간호업무수행에 영향을 미치는 요인을 확인하기 위하여 COVID-19 환자 간호업무수행과 유의한 상관관계를 나타낸 윤리적 민감성, 간호전문직관, 사회적 지지와 일반적 특성 중 유의한 차이를 보였던 가족과 동거 여부를 더미 처리한 후 독립변수로 투입하여 단계적 회귀분석을 시행하였다.

분석 전 회귀분석의 기본가정을 만족하는지 확인하기 위해 P-P 도표와 산점도를 확인한 결과 45° 직선에 근접하여 잔차의 정규성을 만족하였으며, 잔차들이 모두 0을 중심으로 고르게 분포하고 있어 모형의 선형성과 등분산성 가정을 만족하였다. 다중 공선성을 검토한 결과 변수들의 공차한계(tolerance) 값은 0.47~0.56로 0.1 이상이었고, 분산팽창인자(variation inflation factor) 값은 1.80~2.21으로 10을 넘지 않아 독립변수들 간 다중 공선성의 문제는 없었으며, Durbin-Watson 값은 2.17로, 기준 값인 2와 근접하여 잔차들 간의 상관관계가 적어 잔차의 독립성에 관한 검정에서 회귀모형에 적합한 것으로 확인되었다. 회귀분석 결과, COVID-19 환자 간호업무수행에 영향을 미치는 요인은 윤리적 민감성($\beta = .47, p < .001$), 간호전문직관($\beta = .29, p < .001$), 사회적 지지($\beta = .18, p = .014$)로 나타났다. 이들 변수는

감사합니다





03

선형회귀분석

김정아 교수
한양대학교 간호학부

2022년 간호과학연구소 여름방학 통계특강

실습을 중심으로 배우는 알기쉬운 SPSS 중급통계

선형회귀분석

김 정 아

(한양대학교 간호학부 교수)

2022.6.30(목). 13:00-14:40 (100분)

한양대학교 경영관 지하 2층 203호 (PC실)

Linear Regression

사전작업

Data Set 검토

Dummy Variable 만들기

단순회귀분석 다중회귀분석 위계적회귀분석

기본 개념과 연구가설

SPSS로 단순회귀분석하기

SPSS 출력결과의 해석

논문결과표 작성 및 기술

선형회귀분석 실습을 위한 Fake Data Set

일반적 특성 및 직무관련특성	업무환경	기분상태	재직의도
<ul style="list-style-type: none">• 성별, 연령, 결혼상태, 최종학력• 총임상기간, 채용형태, 직위, 근무병동, 병원유형	<ul style="list-style-type: none">• Rossberg, Eiring, & Friis (2004), WES-10 (Working Environment Scale-10• Likert 5 point-Scale• 4개 Subscales<ol style="list-style-type: none">1) 자아실현(4문항)<ul style="list-style-type: none">• 직원들이 지원을 받고 있다고 느끼는 정도, 더 많은 자신감을 얻었는지, 병동에서 일하면서 지식활용이 가능함을 경험 했는 지의 정도.2) 업무량(2문항)<ul style="list-style-type: none">• 직원에게 부과된 과업의 수와 동시에 여러 장소에 존재했어야 한다고 느끼는 정도3) 갈등(2문항)<ul style="list-style-type: none">• 직원들이 직원들 사이에서 갈등과 충성도를 경험하는 정도4) 신경증(2문항)<ul style="list-style-type: none">• 직원들이 출근하는 것에 대해 걱정하는 정도 및 병동에서 얼마나 신경증적이 되는지 혹은 긴장이 되는지	<ul style="list-style-type: none">• McNair et al. (2003), POMS-B• Likert 5-point Scale• 6개 Subscales<ol style="list-style-type: none">1) 긴장 (Tension or Anxiety)2) 우울 (Depression or Dejection)3) 분노 (Anger or Hostility)4) 피로 (Fatigue or Inertia)5) 혼돈 (Confusion or Bewilderment)6) 활기 (Vigor or Activity)	<ul style="list-style-type: none">• Cowin (2002), NRI (Nurses' Retention Index)• 6문항 Likert 8-point Scale

더미변수 만들기

■ 더미변환

- 회귀분석에서 범주형 변수를 통제할 때 사용
- 범주형 자료를 회귀분석에서 활용할 수 있는 형태로 만들어 주는 것

회귀분석을 위한 기본 조건/가정

- 투입되는 독립변수와 종속변수는 모두 (정규분포하는) 연속형변수
- 투입되는 독립변수들은 서로 독립적

범주형 변수는 회귀분석에 투입이 불가능 (범주형 변수는 증가한다 or 감소한다는 개념이 없기 때문)

BUT 우리가 영향력을 궁금해 하는 현실세계의 많은 변수들은 범주형 변수.....어떻게 범주형 변수를 회귀분석에 투입할 수 있을까?

더미변수 만들기

- 더미변환 (cont'd)
 - 회귀분석에서 범주형 변수를 통제할 때 사용
 - 범주형 자료를 회귀분석에서 활용할 수 있는 형태로 만들어 주는 것
- 간호사의 재직의도 영향요인을 확인하려는 연구에 성별을 독립변수로 투입하려고 할 때
 - 성별(1: 남자간호사, 2: 여자간호사)이 높아질 수록 재직의도가 높아진다 or 성별이 낮아질 수록 재직의도가 높아진다 (x)→성별의 더미변수를 만들어 투입해야 함
 - 더미변수의 수는 해당변수 원래 Category의 수-1개
 - 성별: 남,녀 2개 categories→성별 더미변수는 1개
 - 직위: 평간호사, 주임간호사, 수간호사 3개→직위 더미변수는 2개

더미변수 생성의 예시

Category=2→dummy variable 1개

성별	성별 더미변수01
1(남)	0
2(여)	1

Category=3→dummy variable 2개

직위	직위 더미변수01	직위 더미변수02
1(평간호사)	0	0
2(주임간호사)	1	0
3(수간호사)	0	1

Category=4→dummy variable 3개

병원	병원 더미변수01	병원 더미변수02	병원 더미변수 03
1(A 병원)	0	0	0
2(B 병원)	1	0	0
3(C 병원)	0	1	0
4(D 병원)	0	0	1

더미변수 생성의 예시

Category=2→dummy variable 1개

성별	성별 더미변수01
1(남)	0
2(여)	1

Reference (참조항목)

Category=3→dummy variable 2개

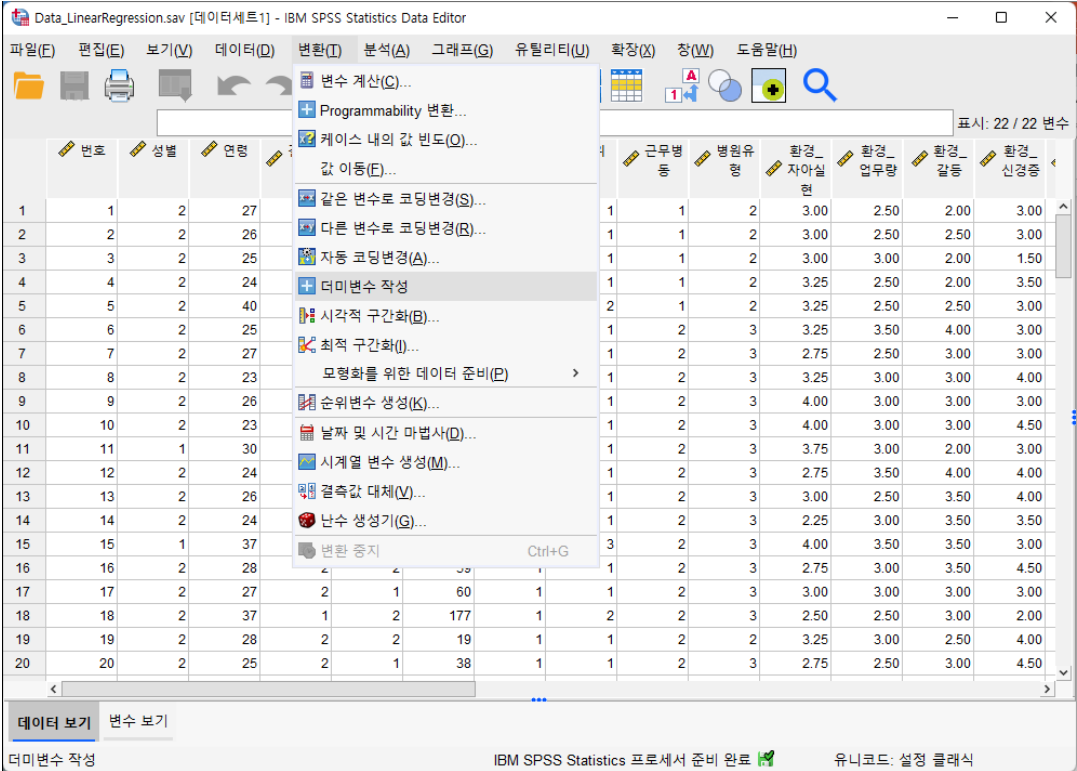
직위	직위 더미변수01	직위 더미변수02
1(평간호사)	0	0
2(주임간호사)	1	0
3(수간호사)	0	1

Category=4→dummy variable 3개

병원	병원 더미변수01	병원 더미변수02	병원 더미변수 03
1(A 병원)	0	0	0
2(B 병원)	1	0	0
3(C 병원)	0	1	0
4(D 병원)	0	0	1

SPSS에서 Dummy Variable 만들기

- 1. 변환>더미변수작성
- 2. 변환>다른변수로 코딩변경



SPSS에서 Dummy Variable 만들기

- 1. 변환>더미변수작성
- 2. 변환>다른변수로 코딩변경

측도 형식의 사전점검 필수!!!!

*Data_LinearRegression.sav [데이터세트1] - IBM SPSS Statistics Data Editor

	이름	유형	너비	소수점이...	레이블	값	결측값	열	맞춤	측도	역할
1	번호	숫자	11	0	번호	지정없음	지정없음	8	오른쪽	척도	입력
2	성별	숫자	11	0		{1, 남자}...	지정없음	8	오른쪽	명목형	입력
3	연령	숫자	11	0	연령(세)	지정없음	지정없음	8	오른쪽	척도	입력
4	결혼상태	숫자	11	0		{1, 기혼}...	지정없음	8	오른쪽	명목형	입력
5	최종학력	숫자	11	0		{1, 전문대학}...	지정없음	8	오른쪽	순서형	입력
6	종업상기간	숫자	8	0	총 임상근무경...	지정없음	지정없음	8	오른쪽	척도	입력
7	재용형태	숫자	11	0	재용형태	{1, 정규직}...	지정없음	8	오른쪽	명목형	입력
8	직위	숫자	11	0	직위	{1, 평간호사}...	지정없음	8	오른쪽	순서형	입력
9	근무병동	숫자	8	0	근무병동	{1, 일반병동}...	지정없음	8	오른쪽	명목형	입력
10	병원유형	숫자	8	0	병원유형_2차 ...	{2, 2차병원}...	지정없음	8	오른쪽	명목형	입력
11	환경_자아...	숫자	8	2	환경-자아실현	지정없음	지정없음	8	오른쪽	척도	입력
12	환경_업무랑	숫자	8	2	환경-업무랑	지정없음	지정없음	8	오른쪽	척도	입력
13	환경_갈등	숫자	8	2	환경-갈등	지정없음	지정없음	8	오른쪽	척도	입력
14	환경_신경증	숫자	8	2	환경-신경질	지정없음	지정없음	8	오른쪽	척도	입력
15	직무만족도	숫자	8	2	총만족도평균평...	지정없음	지정없음	8	오른쪽	척도	입력
16	긴장	숫자	8	2	기본상태_긴장	지정없음	지정없음	8	오른쪽	척도	입력
17	우울	숫자	8	2	기본상태_우울	지정없음	지정없음	8	오른쪽	척도	입력
18	분노	숫자	8	2	기본상태_분노	지정없음	지정없음	8	오른쪽	척도	입력
19	활기	숫자	8	2	기본상태_활기	지정없음	지정없음	8	오른쪽	척도	입력
20	피로	숫자	8	2	기본상태_피로	지정없음	지정없음	8	오른쪽	척도	입력
21	혼돈	숫자	8	2	기본상태_혼돈	지정없음	지정없음	8	오른쪽	척도	입력
22	재직의도	숫자	8	2	총재직의도총점	지정없음	지정없음	8	오른쪽	척도	입력
23	성별_더미_1	숫자	8	2	성별=남자	지정없음	지정없음	17	오른쪽	명목형	입력
24	성별_더미_2	숫자	8	2	성별=여자	지정없음	지정없음	17	오른쪽	명목형	입력

사전작업

DataSet 검토

Dummy Variable 만들기

단순회귀분석

기본 개념과 연구가설

SPSS로 분석하기

SPSS 출력결과표 해석

논문결과표 작성 및 기술

다중회귀분석

위계적 회귀분석

변환(T) 분석(A) 그래프(G) 유틸리티(U)

변수 계산(C)...

Programmability 변환...

케이스 내의 값 빈도(O)...
값 이동(E)...

같은 변수로 코딩변경(S)...

다른 변수로 코딩변경(R)...

자동 코딩변경(A)...

더미변수 작성

시각적 구간화(B)...

최적 구간화(I)...

모형화를 위한 데이터 준비(P) >

순위변수 생성(K)...

날짜 및 시간 마법사(D)...

시계열 변수 생성(M)...

결측값 대체(V)...

난수 생성기(G)...

변환 중지 Ctrl+G

더미변수 작성

변수(I):
번호 [번호]
연령(세) [연령]
결혼상태
최종학력
총 임상근무경력(개월) [총임상기간]
채용형태 [채용형태]
직위 [직위]
근무병동 [근무병동]
변환 유형 선택 및 코딩 방법 선택 (변환 유형)

더미변수 레이블

☒ 값 레이블 사용(V)
☐ 값 사용(S)

값 순서

☒ 오름차순(A)
☐ 내림차순(D)

매크로

☐ 매크로 정의에서 첫 번째 더미 범주 생략(E)
참고: 매크로 이름은 일반적으로 !로 시작됩니다.

측정 수준 사용

☒ 척도변수 값에 대한 더미를 작성하지 않음(O)
☐ 모든 변수에 대한 더미 작성(C)

다음에 대한 더미변수 작성(C):
성별

주효과 더미변수
☒ 주효과 더미 작성(M)
루트 이름(선택한 변수당 한 개)(Q):
성별_더미
매크로 이름(N):

이원 상호작용
☐ 모든 이원 상호작용에 대한 더미 작성(I)
루트 이름(E):
매크로 이름(C):

삼원 상호작용
☐ 모든 삼원 상호작용에 대한 더미 작성(H)
루트 이름:
매크로 이름:

이 대화 상자에는 Python Essentials가 필요합니다.

확인

불러넣기(P)

재설정(R)

취소

도움말

Reference (참조항목)을 제거

*Data_LinearRegression.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(E) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)

2 :

	성별_더미_1	성별_더미_2	변수	변수
1	.00	1.00		
2	.00	1.00		
3	.00	1.00		
4	.00	1.00		
5	.00	1.00		
6	.00	1.00		
7	.00	1.00		
8	.00	1.00		
9	.00	1.00		
10	.00	1.00		
11	1.00	.00		
12	.00	1.00		
13	.00	1.00		
14	.00	1.00		
15	1.00	.00		
16	.00	1.00		
17	.00	1.00		
18	.00	1.00		
19	.00	1.00		
20	.00	1.00		
21	1.00	.00		
22	.00	1.00		
23	.00	1.00		

데이터 보기 변수 보기

DataSet에서 더미변수 변환

변수	측도	응답범주	더미변수 수	참조항목
성별	명목형	1=남자, 2=여자	1	남자
결혼상태	명목형	1=기혼, 2=미혼	1	기혼
최종학력	순서형	1=전문대졸, 2=대졸	1	전문대졸
채용형태	명목형	1=정규직, 2=비정규직	1	정규직
직위	순서형	1=평간호사, 2=책임간호사, 3=수간호사	2	평간호사
근무병동	명목형	1=일반병동, 2=특수병동(ICU, OR, RR, ER))	1	일반병동
병원유형	명목형	2=2차병원, 3=3차병원	1	2차병원

DataSet에서 더미변수 변환

IBM SPSS Statistics Data Editor

	이름	유형	너비	소수점이...	레이블	값	결측값	열	맞춤	측!
5	최종학력	숫자	11	0		{1, 전문대학...	지정없음	8	오른쪽	순서형
6	총임상기간	숫자	8	0	총 임상근무경력(개월)	지정없음	지정없음	8	오른쪽	척도
7	채용형태	숫자	11	0	채용형태	{1, 정규직}...	지정없음	8	오른쪽	명목형
8	직위	숫자	11	0	직위	{1, 평간호사...	지정없음	8	오른쪽	순서형
9	근무병동	숫자	8	0	근무병동	{1, 일반병동...	지정없음	8	오른쪽	명목형
10	병원유형	숫자	8	0	병원유형_2차 및 3차 병원 여부	{2, 2차병원}...	지정없음	8	오른쪽	명목형
11	환경_자아실현	숫자	8	2	환경-자아실현	지정없음	지정없음	8	오른쪽	척도
12	환경_업무량	숫자	8	2	환경-업무량	지정없음	지정없음	8	오른쪽	척도
13	환경_갈등	숫자	8	2	환경-갈등	지정없음	지정없음	8	오른쪽	척도
14	환경_신경증	숫자	8	2	환경-신경질	지정없음	지정없음	8	오른쪽	척도
15	직무만족도	숫자	8	2	충만족도평균평점	지정없음	지정없음	8	오른쪽	척도
16	긴장	숫자	8	2	기본상태_긴장	지정없음	지정없음	8	오른쪽	척도
17	우울	숫자	8	2	기본상태_우울	지정없음	지정없음	8	오른쪽	척도
18	분노	숫자	8	2	기본상태_분노	지정없음	지정없음	8	오른쪽	척도
19	활기	숫자	8	2	기본상태_활기	지정없음	지정없음	8	오른쪽	척도
20	피로	숫자	8	2	기본상태_피로	지정없음	지정없음	8	오른쪽	척도
21	혼돈	숫자	8	2	기본상태_혼돈	지정없음	지정없음	8	오른쪽	척도
22	재직의도	숫자	8	2	충재직의도충전	지정없음	지정없음	8	오른쪽	척도
23	성별_더미_2	숫자	8	2	성별=여자	지정없음	지정없음	17	오른쪽	명목형
24	결혼상태_더미_2	숫자	8	2	결혼상태=미혼	지정없음	지정없음	23	오른쪽	명목형
25	최종학력_더미_2	숫자	8	2	최종학력=대학졸업	지정없음	지정없음	23	오른쪽	명목형
26	채용형태_더미_2	숫자	8	2	채용형태=비정규직	지정없음	지정없음	23	오른쪽	명목형
27	직위_더미_2	숫자	8	2	직위=책임간호사	지정없음	지정없음	17	오른쪽	명목형
28	직위_더미_3	숫자	8	2	직위=수간호사	지정없음	지정없음	17	오른쪽	명목형
29	근무병동_더미_2	숫자	8	2	근무병동=특수병동(ICU, OR, ...)	지정없음	지정없음	23	오른쪽	명목형
30	병원유형_더미_2	숫자	8	2	병원유형=3차 병원	지정없음	지정없음	23	오른쪽	명목형

생성된 (Reference (참조항목) 제거 이후)
더미변수 확인

기본개념과 연구가설

■ Simple Linear Regression Analysis

- 연속형 독립변수가 연속형 종속변수에 미치는 영향을 검증하는 분석방법 → 연속형 독립변수가 증가할 때, 연속형 종속변수는 **증가**하는지 or **감소**하는지 확인
 - 비례 → 정(+)의 영향
 - 반비례 → 부(-)의 영향

■ 연구가설

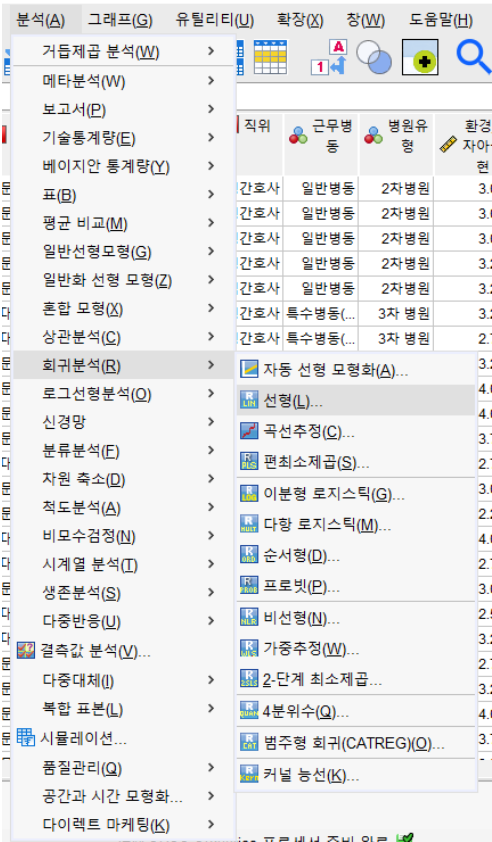
- 독립변수가 종속변수에 영향을 미칠 것이다.
- 독립변수가 종속변수에 유의한 정(+)의 영향을 미칠 것이다.
- 독립변수가 종속변수에 유의한 부(-)의 영향을 미칠 것이다.

기본개념과 연구가설

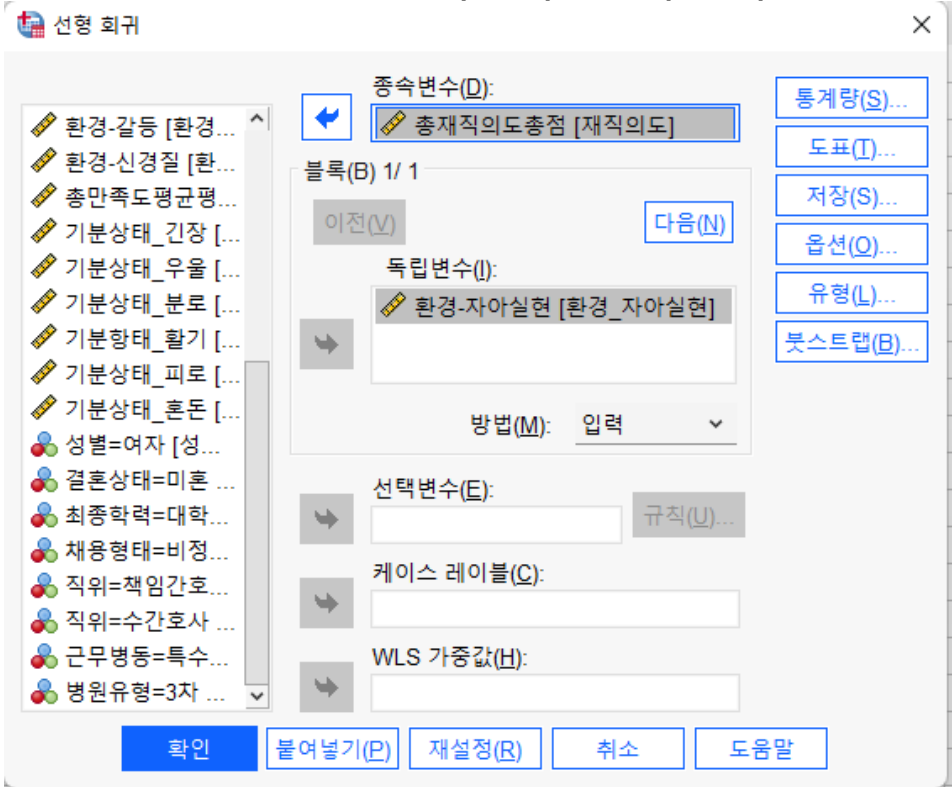
■ 연구가설_실습 예시

- 연구문제
 - 간호사가 인지하는 업무환경_자아실현은 간호사의 재직의도에 영향을 미칠까?
- 연구가설
 - 간호사가 인지하는 업무환경_자아실현은 간호사의 재직의도에 유의한 영향을 미칠 것이다.
 - 간호사가 인지하는 업무환경_자아실현은 간호사의 재직의도에 유의한 정(+)의 영향을 미칠 것이다.
 - 간호사가 인지하는 업무환경_자아실현은 간호사의 재직의도에 유의한 부(-)의 영향을 미칠 것이다.

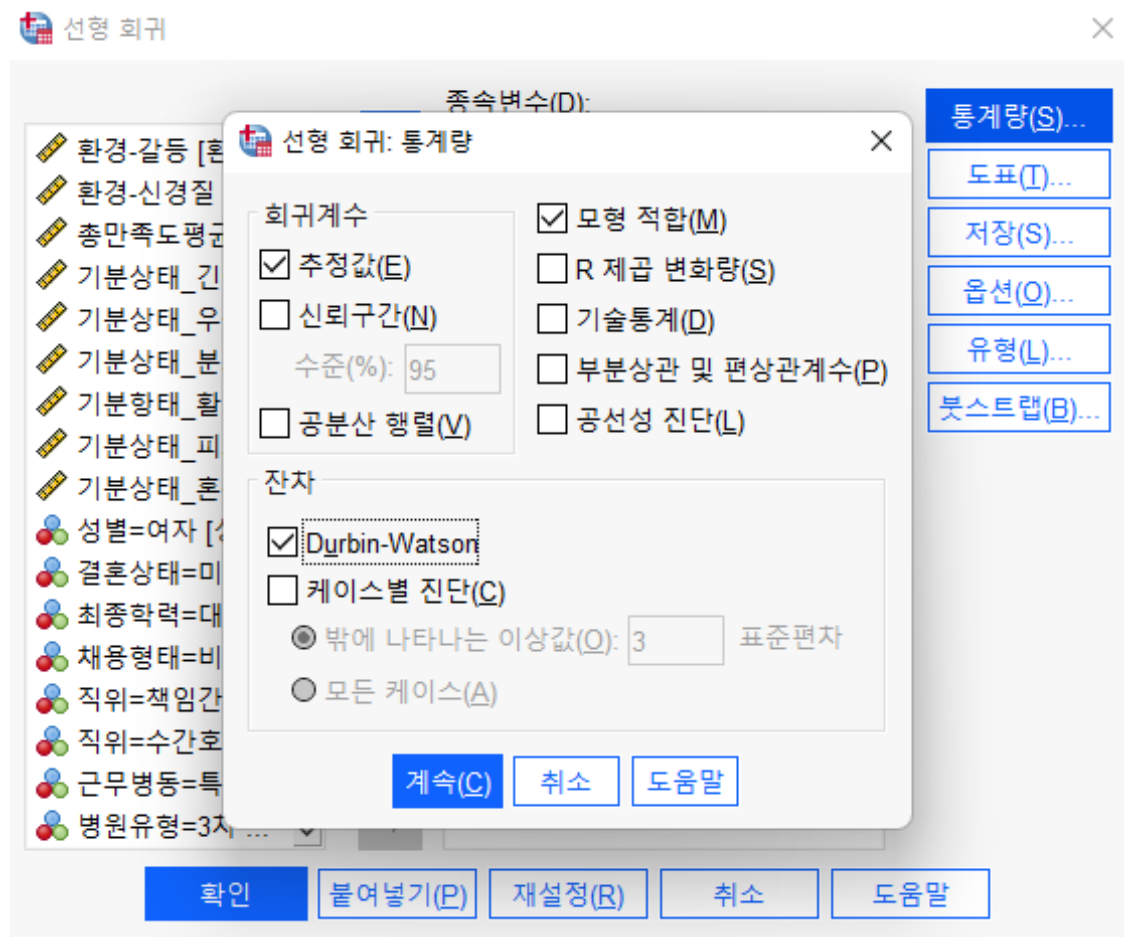
분석>회귀분석>선형



선형회귀 창에서 독립변수와
종속변수 클릭 &이동



통계량 클릭→통계량 창에서 Durbin-Watson 클릭→계속→확인



1. (독립변수의 유의성 여부 확인 전에) 회귀모형의 적합도 및 설명력 확인
- 적합도: ANOVA table
 - 설명력: 모형요약의 R제곱 값 확인

R제곱과 계산방식이 약간 다른 설명력 수치→단순회귀분석에서는 큰 의미가 없고, 다중회귀분석에서 그 의미가 중요

잔차의 독립성 여부를 판단: 잔차는 어떤 규칙 or 경향이 없이 랜덤한가?
잔차=관측값-예측값→회귀분석에서 오차의 개념
Durbin-Watson값이 2에 근사할 수록 잔차에 독립성이 있음
1.5~2.5 or 1-3 등 독립성 판단기준은 학자에 따라 일부 차이

모형 요약^b

모형	R	R 제곱	수정된 R 제곱	추정값의 표준 오차	Durbin-Watson
1	.249 ^a	.062	.053	7.40430	1.475

a. 예측자: (상수), 환경-자아실현
b. 종속변수: 총재직의도총점

설명력은 6.2%→”간호사가 인지한 업무환경_자아실현은 간호사의 재직의도를 6.2% 설명한다.”

ANOVA^a

모형		제곱합	자유도	평균제곱	F	유의확률
1	회귀	363.805	1	363.805	6.636	.011 ^b
	잔차	5482.362	100	54.824		
	전체	5846.167	101			

a. 종속변수: 총재직의도총점
b. 예측자: (상수), 환경-자아실현

F값은 6.636, p값은 .011→”회귀모형이 적합”

- 2. 회귀계수의 유의성 확인
- 3. 회귀계수가 정(+)적으로 유의한지 or 부(-)적으로 유의한지 확인

계수 ^a						
		비표준화 계수		표준화 계수		
모형		B	표준화 오류	베타	t	유의확률
1	(상수)	20.051	4.760		4.213	<.001
	환경-자아실현	3.811	1.479	.249	2.576	.011

a. 종속변수: 총재직의도 총점

환경_자아실현의 p값은 .011로 유의함→간호사가 인지하는 업무환경_자아실현은 간호사의 재직의도에 유의한 영향을 미침

비표준화 계수가 3.811, 표준화 계수가 .249로 모두 양(+)의 값→환경_자아실현은 재직의도에 정(+)의 영향을 미침→간호사가 인지하는 업무환경_자아실현은 간호사의 재직의도에 유의한 정(+)의 영향을 미침

참고) 비표준화 계수와 표준화 계수의 부호는 서로 동일함→단순회귀분석에서는 영향력의 방향(정 or 부)의 판단을 위해서는 두 계수 중 어느 값을 보아도 무관→다만, 다중회귀분석에서는 표준화 계수와 비표준화 계수를 구분해야 함

파일>내보내기 기능을 활용한 논문결과표 작성

파일(E) 편집(E) 보기(V) 데이터(D) 변환(T) 삽입(I) 형식(O) 분석(A) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)

새로 만들기(N) >
열기(O) >
데이터 가져오기(D) >
일반 열기
복원점 열기(P)...
닫기(C) Ctrl+F4
저장(S) Ctrl+S
다른 이름으로 저장(A)...
복원점 저장(S)...
내보내기(T)...
파일에 자동 복구 사용(E)
데이터 파일 정보 표시(I) >
프로세서 중단 Ctrl+Period
변수 정보 수집
페이지 속성(G)...
페이지 설정(U)...
인쇄 미리보기(V)
인쇄(P)... Ctrl+P
시작 대화 상자(W)...
최근에 사용한 데이터(Y) >
최근에 사용한 파일(E) >
사용권 관리(G)
로그아웃 및 종료(G)
종료(X)

변수 작성

레이블

미_1	병원유형=2차 병원
미_2	병원유형=3차 병원

C:\Users\User\Dropbox\Lecture\2022_1\간호과학연구소 통계특강_회귀분석\Data n Analysis for 선형회귀

입력/제거된 변수^a

입력된 변수	제거된 변수	방법
경-자아실현 ^b		입력

수: 총재직의도총점
모든 변수가 입력되었습니다.

모형 요약^b

R	R 제곱	수정된 R 제곱	추정값의 표준 오차	Durbin-Watson
.249 ^a	.062	.053	7.40430	1.475

(상수), 환경-자아실현
수: 총재직의도총점

ANOVA^a

내보내기(T)...

내보내기 출력결과

내보낼 개체

☒ 모두(A) ☐ 모두 표시(V) ☐ 선택(D)

문서

유형(I):

Excel 2007 이상(*.xlsx)

Excel 97-2004(*.xls)
Excel 2007 이상(*.xlsx)
Excel 2007 이상 Macro 활성화됨(*.xlsm)
HTML(*.htm)
Portable Document Format(*.pdf)
PowerPoint(*.ppt)
텍스트 - 일반(*.txt)
텍스트 - UTF8(*.txt)

옵션(O):

원하는 작업	워크북 만들기
워크시트 이름	
워크시트의 위치	마지막 열 다음에
피벗표의 레이어	인쇄 레이어 설정에 유효(표를 ...
각주와 캡션 포함	예
모형 보기	인쇄 설정에 유효(각 모형에 대...

옵션 변경(C)...

파일 이름(E):

user\Dropbox\Lecture\2022_1\간호과학연구소 통계특강_회귀분석\단순회귀분석 결과표.xlsx

찾아보기(B)...

그래픽

유형(E):

JPEG 파일(*.jpg)

옵션(S):

사용 가능한 옵션이 없음

옵션 변경(H)...

루트 파일 이름(N):

C:\Users\user\Dropbox\내 PC (DESKTOP-UC0U9NR)\Documents\출력.jpeg

찾아보기(W)...

☐ 포함 폴더 열기(O)

확인

붙여넣기(P)

재설정(R)

취소

도움말

성

23

파일>내보내기 기능을 활용한 논문결과표 작성

자동 저장 100% 단순회귀분석 결과표

검색(Alt+Q)

파일 홈 삽입 페이지 레이아웃 수식 데이터 검토 보기 도움말

기본값

유지

나가기

새로 만들기

옵션

기본

페이지 나누기

페이지 미리 보기

레이아웃

사용자 지정 보기

논금자

수식 입력줄

논금선

머리글

확대/축소

100%

선택 영역 확대/축소

새 창

모두 정렬

틀 고정

숨기기

숨기기 취소

창

시트 보기

통합 문서 보기

표시

확대/축소

J498

fx

파일>내보내기 기능을 활용한 논문결과표 작성

[표 1] 업무환경_자아실현이 재직의도에 미치는 영향

[N=102]

종속변수	독립변수	B	SE	β	t	p
재직의도	(상수)	20.051	4.760		4.213	<.001
	업무환경_자아실현	3.811	1.479	0.249	2.576	0.011

$F=6.636$ ($p=.011$), $R^2=.062$, $D-W=1.475$

간호사가 인지한 업무환경의 자아실현 하위영역이 재직의도에 미치는 영향을 검증하기 위해 단순회귀분석을 시행한 결과가 [표 1]에 제시되어 있다.

회귀모형은 통계적으로 유의하였으며($F=6.636$, $p=.011$), 회귀모형의 설명력은 약 6.2%로 나타났다. Durbin-Watson 통계량은 1.475로 나타나 (2에 근사한 값을 보여) 잔차의 독립성 가정에 위배되지 않는 것으로 나타났다 (잔차의 독립성 가정에 문제가 없는 것으로 나타났다).

회귀계수의 유의성 검증결과, 간호사가 인지한 업무환경의 자아실현 하위영역은 재직의도에 유의한 정(+)의 영향을 미치는 것으로 나타났다($t=2.576$, $p=.011$). 즉, 간호사가 인지하는 업무환경의 자아실현 정도가 높아질 수록 간호사의 재직의도도 높아지는 것으로 평가되었다.

기본개념

■ Multiple Linear Regression Analysis

- 두개 이상의 연속형 독립변수가 연속형 종속변수에 미치는 영향을 검증하는 분석방법→독립변수 중 어떤 변수가 종속변수에 영향을 미치는지 & 해당 연속형 독립변수가 증가할 때, 연속형 종속변수는 **증가**하는지 or **감소**하는지 확인
 - 비례→정(+)의 영향
 - 반비례→부(-)의 영향

■ 연구가설

- 독립변수가 종속변수에 영향을 미칠 것이다.
- 독립변수가 종속변수에 유의한 정(+)의 영향을 미칠 것이다.
- 독립변수가 종속변수에 유의한 부(-)의 영향을 미칠 것이다.

연구가설

■ 연구가설

- 독립변수 01이 종속변수에 유의한 영향을 미칠 것이다.
- 독립변수 02가 종속변수에 유의한 영향을 미칠 것이다.
- \vdots
- 독립변수 n이 종속변수에 유의한 영향을 미칠 것이다.

연구가설

■ 연구가설 (cont'd)

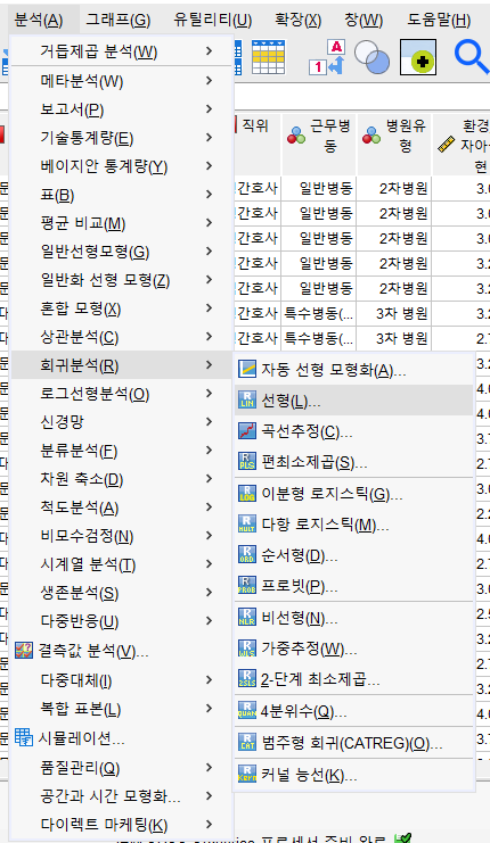
- 독립변수 01이 종속변수에 유의한 정(+)의 영향을 미칠 것이다.
- 독립변수 02가 종속변수에 유의한 정(+)의 영향을 미칠 것이다.
- \vdots
- 독립변수 n이 종속변수에 유의한 정(+)의 영향을 미칠 것이다.

사전작업
DataSet 검토
Dummy Variable 만들기
단순회귀분석
다중회귀분석
기본 개념과 연구가설
SPSS로 분석하기
SPSS 출력결과표 해석
논문결과표 작성 및 기술
위계적 회귀분석

연구가설

- 연구가설 (cont'd)
 - 독립변수 01이 종속변수에 유의한 음(-)의 영향을 미칠 것이다.
 - 독립변수 02가 종속변수에 유의한 음(-)의 영향을 미칠 것이다.
 - ⋮
 - 독립변수n이 종속변수에 유의한 음(-)의 영향을 미칠 것이다.

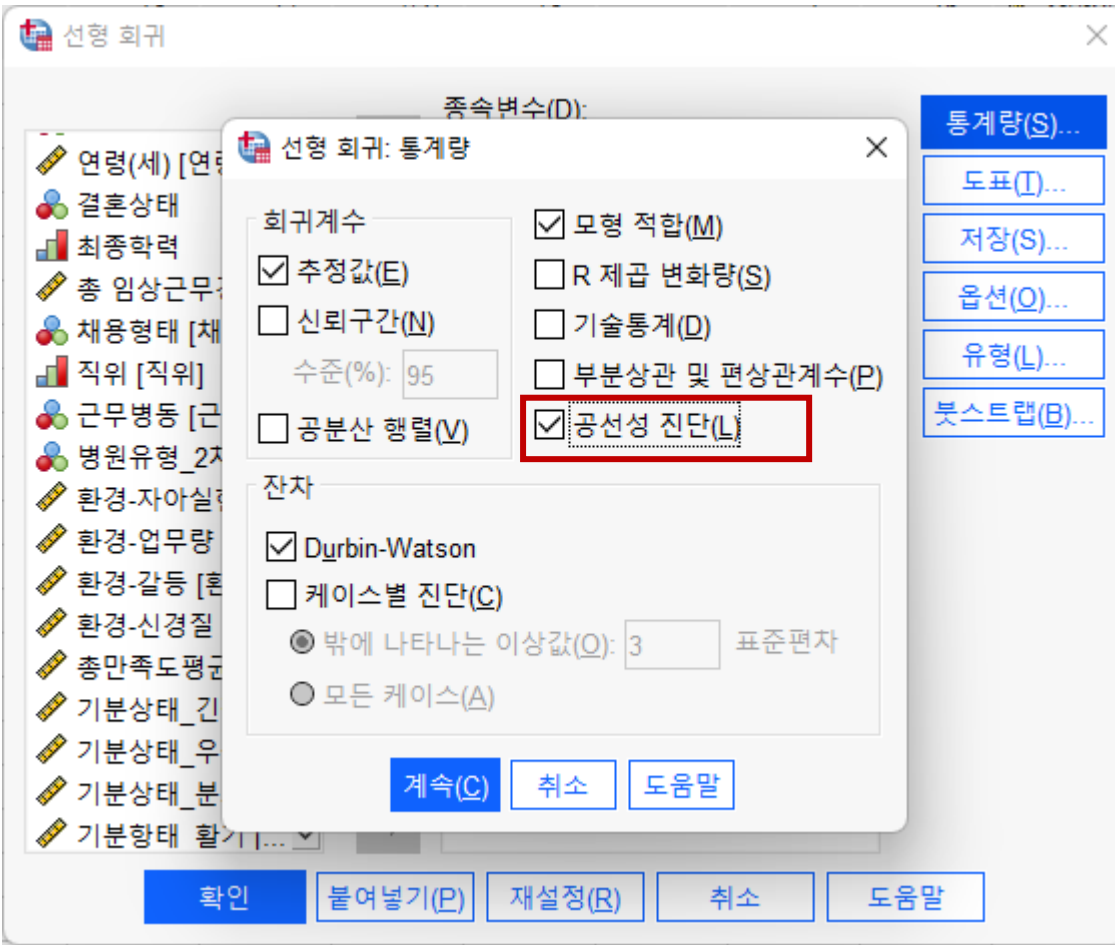
분석>회귀분석>선형



선형회귀 창에서 독립변수와
종속변수 클릭 &이동



통계량 클릭→통계량 창에서 Durbin-Watson 클릭
→공선성진단 클릭→계속→확인



1. (독립변수의 유의성 여부 확인 전에) 회귀모형의 적합도 및 설명력 확인
- 적합도: ANOVA table
 - 설명력: 모형요약의 R제곱 값 확인

수정된 R제곱: 변수의 개수를 분모에 포함시켜 계산→불필요한 독립변수가 추가될 경우 수정된 R제곱값이 감소→회귀모형에 불필요한 변수를 판단하기 유리한 설명력 수치

잔차의 독립성 여부를 판단: 잔차는 어떤 규칙 or 경향이 없이 랜덤한가?
잔차=관측값-예측값→회귀분석에서 오차의 개념
Durbin-Watson값이 2에 근사할 수록 잔차에 독립성이 있음
1.5~2.5 or 1-3 등 독립성 판단기준은 학자에 따라 일부 차이

모형 요약^b

모형	R	R 제곱	수정된 R 제곱	추정값의 표준 오차	Durbin-Watson
1	.561 ^a	.315	.240	6.63465	1.746

- a. 예측자: (상수), 기분상태_혼돈, 환경-자아실현, 환경-업무량, 환경-신경질, 기분향태_활기, 환경-갈등, 기분상태_우울, 기분상태_분로, 기분상태_피로, 기분상태_긴장
- b. 종속변수: 총재직의도총점

설명력은 R제곱은 31.5%, 수정된 R제곱은 24.0%
다중회귀분석에서는 R제곱과 수정된 R제곱 값을 함께 표시 or 수정된 R 제곱값을 표시

ANOVA^a

모형		제곱합	자유도	평균제곱	F	유의확률
1	회귀	1840.477	10	184.048	4.181	<.001 ^b
	잔차	4005.690	91	44.019		
	전체	5846.167	101			

- a. 종속변수: 총재직의도총점
- b. 예측자: (상수), 기분상태_혼돈, 환경-자아실현, 환경-업무량, 환경-신경질, 기분향태_활기, 환경-갈등, 기분상태_우울, 기분상태_분로, 기분상태_피로, 기분상태_긴장

F값은 4.181, p값은 <.001→"회귀모형이 적합"

1. (독립변수의 유의성 여부 확인 전에) 회귀모형의 적합도 및 설명력 확인
- 적합도: ANOVA table
 - 설명력: 모형요약의 R제곱 값 확인

수정된 R제곱: 변수의 개수를 분모에 포함시켜 계산→불필요한 독립변수가 추가될 경우 수정된 R제곱값이 감소→회귀모형에 불필요한 변수를

잔차의 독립성 여부를 판단: 잔차는 어떤 규칙 or 경향이 없이 랜덤한가?
잔차=관측값-예측값→회귀분석에서 오차의 개념
Durbin-Watson값이 2에 근사할 수록 잔차에 독립성이 있음
1.5~2.5 or 1-3 등 독립성 판단기준은 학자에 따라 일부 차이

R제곱 (R²)	수정된 R제곱 (adj. R²)
<ul style="list-style-type: none">▪ 불필요한 독립변수가 추가되어도 감소하지 않음▪ 단순회귀분석, 다중회귀분석에서 모두 표기▪ R2, R-Square 등으로 표기	<ul style="list-style-type: none">▪ 불필요한 독립변수가 추가되면 감소함▪ 다중회귀분석에서만 표기▪ adj. R2, adj. R-square 등으로 표기

모형		제곱합	자유도	평균제곱	F	유의확률
1	회귀	1840.477	10	184.048	4.181	<.001 ^b
	잔차	4005.690	91	44.019		
	전체	5846.167	101			

- a. 종속변수: 총재직의도총점
- b. 예측자: (상수), 기분상태_혼돈, 환경-자아실현, 환경-업무량, 환경-신경질, 기분상태_활기, 환경-갈등, 기분상태_우울, 기분상태_분로, 기분상태_피로, 기분상태_긴장

F값은 4.181, p값은 <.001→"회귀모형이 적합"

- 2. 회귀계수의 유의성과 다중공선성 확인
- 3. 회귀계수가 정(+)적으로 유의한지 or 부(-)적으로 유의한지 확인

계수 ^a								
두 번 눌러서 활성화		비표준화 계수		표준화 계수	t	유의확률	공선성 공차	통계량 VIF
		B	표준화 오류	베타				
1	(상수)	29.284	8.624		3.396	.001		
	환경-자아실현	.719	1.540	.047	.467	.642	.741	1.350
	환경-업무량	.196	2.263	.008	.086	.931	.847	1.181
	환경-갈등	-1.616	1.446	-.122	-1.117	.267	.628	1.591
	환경-신경질	1.089	1.190	.094	.915	.363	.713	1.403
	기분상태_긴장	-.212	.343	-.102	-.619	.538	.277	3.612
	기분상태_우울	-.420	.200	-.249	-2.097	.039	.536	1.866
	기분상태_분로	-.246	.326	-.113	-.756	.451	.338	2.956
	기분상태_활기	.679	.251	.300	2.703	.008	.610	1.638
	기분상태_피로	-.056	.249	-.034	-.223	.824	.333	3.007
	기분상태_혼돈	.232	.388	.103	.597	.552	.254	3.930

a. 종속변수: 총재직의도총점

- 2. 기분상태_우울의 p값은 .039로, 기분상태_활기의 p값은 .008로 유의함→간호사의 기분상태 하위영역 중 우울과 활기가 간호사의 재직의도에 유의한 영향을 미침
- 3. 기분상태_우울의 비표준화 계수가 -.420, 표준화 계수가 -.249로 모두 양(-)의 값→기분상태_우울은 재직의도에 부(-)의 영향을 미침→간호사의 기분상태 중 우울은 간호사의 재직의도에 유의한 부(-)의 영향을 미침
- 4. 기분상태_활기의 비표준화 계수가 .679, 표준화 계수가 .300으로 간호사의 기분상태 중 활기는 간호사의 재직의도에 유의한 정(+)의 영향을 미침→간호사의 기분상태 중 활기는 간호사의 재직의도에 유의한 정(+)의 영향을 미침
- 1. 다중공선성 확인: 분산팽창지수(Variance Inflation Factor, VIF) 값으로 판단→10 미만이면 다중공선성에 문제가 없다고 판단 or 10미만이라도 5를 초과한다면 다중공선성을 의심해 볼 수 있음
 - 다중공선성
 - ✓ 독립변수간 유사성 의미→독립변수끼리 유사성이 너무 높으면 서로의 영향력을 감소시킬 수 있음. 즉, 너무 비슷한 변수 n개가 동시에 투입되면서 그 영향력이 불필요하게 나뉘는 것→유의하게 나올 수 있는 변수가 유의하지 않게 나올 수 있다는 것

- 2. 회귀계수의 유의성과 다중공선성 확인
- 3. 회귀계수가 정(+)적으로 유의한지 or 부(-)적으로 유의한지 확인

		계수 ^a						통계량 VIF
두 번 눌러서 활성화		비표준화 계수 B 표준화 오류		표준화 계수 베타	t	유의확률	공차	
1	(상수)	29.284	8.624		3.396	.001		
	환경-자아실현	.719	1.540	.047	.467	.642	.741	1.350
	환경-업무량	.196	2.263	.008	.086	.931	.847	1.181
	환경-갈등	-1.616	1.446	-.122	-1.117	.267	.628	1.591
	환경-신경질	1.089	1.190	.094	.915	.363	.713	1.403
	기분상태_긴장	-.212	.343	-.102	-.619	.538	.277	3.612
	기분상태_우울	-.420	.200	-.249	-2.097	.039	.536	1.866
	기분상태_분로	-.246	.326	-.113	-.756	.451	.338	2.956
	기분상태_활기	.679	.251	.300	2.703	.008	.610	1.638
	기분상태_피로	-.056	.249	-.034	-.223	.824	.333	3.007
	기분상태_혼돈	.232	.388	.103	.597	.552	.254	3.930

a. 종속변수: 총재직의도 총점

비표준화 계수	표준화 계수
<ul style="list-style-type: none">단위가 통일되지 않음절대적인 영향력의 크기변수끼리 영향력 크기를 비교할 수 없음회귀식에 사용되는 계수(독립변수 1점 증가 시 종속변수 증가량)	<ul style="list-style-type: none">단위가 통일됨상대적인 영향력의 크기변수끼리 영향력 크기를 비교할 수 있음

여러 독립변수의 상대적인 영향을 비교하는 다중회귀분석에서는 표준화계수를 우선적으로 고려→유의한 영향력을 가지는 것으로 나타난 독립변수의 표준화 계수 값을 읽어줌

[표 2] 업무환경_자아실현이 재직의도에 미치는 영향

								(N=102)
종속변수	독립변수	B	SE	β	t	p	VIF	
다중회귀분석	(상수)	29.284	8.624		3.396	0.001		
	업무환경_자아실현	0.719	1.540	0.047	0.467	0.642	1.350	
	업무환경_업무량	0.196	2.263	0.008	0.086	0.931	1.181	
	업무환경_갈등	-1.616	1.446	-0.122	-1.117	0.267	1.591	
	업무환경_신경증	1.089	1.190	0.094	0.915	0.363	1.403	
	재직의도_기분상태_긴장	-0.212	0.343	-0.102	-0.619	0.538	3.612	
	기분상태_우울	-0.420	0.200	-0.249	-2.097	0.039	1.866	
	기분상태_분노	-0.246	0.326	-0.113	-0.756	0.451	2.956	
	기분상태_활기	0.679	0.251	0.300	2.703	0.008	1.638	
	기분상태_피로	-0.056	0.249	-0.034	-0.223	0.824	3.007	
	기분상태_혼돈	0.232	0.388	0.103	0.597	0.552	3.930	
$F=4.181$ ($p<.001$), $R^2=.315$, $adj\ R^2=.240$, $D-W=1.746$								

간호사가 인지한 업무환경과 기분상태의 각 하위영역이 재직의도에 미치는 영향을 검증하기 위해 다중회귀분석을 시행한 결과가 [표 2]에 제시되어 있다.

회귀모형은 통계적으로 유의하였으며($F=4.181$, $p<.001$), 회귀모형의 설명력은 약 24.7%로 나타났다. Durbin-Watson 통계량은 1.746으로 2에 근사한 값을 보여 잔차의 독립성 가정에 문제가 없는 것으로 평가되었고, 분산팽창지수(Variance Inflation Factor, VIF)도 모두 10미만으로 나타나 다중공선성 문제는 없는 것으로 판단되었다.

회귀계수의 유의성 검증결과, 기분상태 하위영역 중 우울($\beta=-.249$, $p=.039$)이 간호사의 재직의도에 유의한 부(-)의 영향을, 기분상태 하위영역 중 활기 ($\beta=.300$, $p=.008$)가 간호사의 재직의도에 유의한 정(+)의 영향을 미치는 것으로 나타났다. 즉, 간호사가 우울할 수록 재직의도는 감소하며, 활기가 증가할 수록 재직의도도 높아지는 것으로 평가되었으며, 표준화계수의 크기를 비교할 때 기분상태 하위영역의 활기가 재직의도에 가장 큰 영향을 그 다음으로 우울이 영향을 미치는 것으로 검증되었다.

기본개념

- Hierarchical Regression Analysis
 - 변수를 추가해가면서 단계적으로 진행하는 회귀분석
 - 회귀분석을 여러 번 진행하는 것
 - (일반적으로) 적은 수의 독립변수만 투입→그 다음 단계에서 추가로 독립변수를 투입→또 다음단계에서 더 많은 독립변수를 투입
 - (일반적으로) 일반적 특성 중 단변량분석에서 유의한 것으로 확인된 변수를 투입: 통제변수→독립변수를 투입(2단계 이후 + α 단계)
 - (일반적으로) 매개효과와 조절효과를 검증할 때 많이 활용되는 방법
 - 목적
 - 각 단계별로 설명력이 어떻게 변화하는지를 확인

연구가설

- 다중회귀분석의 연구가설과 동일

단변량 분석 결과

변수	측도	응답범주	검증방법	유의성
성별	명목형	1=남자, 2=여자	t-test	Non-sig
결혼상태	명목형	1=기혼, 2=미혼	t-test	기혼>미혼
최종학력	순서형	1=전문대졸, 2=대졸	t-test	Non-sig
채용형태	명목형	1=정규직, 2=비정규직	t-test	정규직>비정규직
직위	순서형	1=평간호사, 2=책임간호사, 3=수간호사	Oneway ANOVA	Non-sig
근무병동	명목형	1=일반병동, 2=특수병동 (ICU, OR, RR, ER))	t-test	일반병동>특수병동
병원유형	명목형	2=2차병원, 3=3차병원	t-test	2차병원>3차병원

- 1단계: 일반적특성 중 단변량분석에서 유의했던 변수의 투입
- 2단계: 업무환경 하위영역 추가투입
- 3단계: 기분상태 하위영역 추가투입

선행 회귀

환경-갈등 [환경...]

환경-신경질 [환...]

총만족도평균평...

기분상태_긴장 [...]

기분상태_우울 [...]

기분상태_분로 [...]

기분상태_활기 [...]

기분상태_피로 [...]

기분상태_혼돈 [...]

성별=여자 [성...]

결혼상태=미혼 ...

최종학력=대학...

채용형태=비정...

직위=책임간호...

직위=수간호사 ...

근무병동=특수...

병원유형=3차 ...

종속변수(D):

총재직의도총점 [재직의도]

블록(B) 1/ 1

이전(V)

다음(N)

블록(B) 1/ 1

결혼상태=미혼 [결혼상태_...]

채용형태=비정규직 [채용형...]

근무병동=특수병동/ICU ...

방법(M): 입력

선택변수(E):

규칙(U)...

케이스 레이블(C):

WLS 가중값(H):

통계량(S)...

도표(I)...

저장(S)...

옵션(O)...

유형(L)...

붓스트랩(B)...

확인

붙여넣기(P)

재설정(R)

취소

도움말

- 1단계: 일반적특성 중 단변량분석에서 유의했던 변수의 투입
- 2단계: 업무환경 하위영역 추가투입
- 3단계: 기분상태 하위영역 추가투입

선행 회귀

직위 [직위]

근무병동 [근무...]

병원유형_2차 ...

환경-자아실현 [...]

환경-업무량 [환...

환경-갈등 [환경...

환경-신경질 [환...

총만족도평균평...

기분상태_긴장 [...]

기분상태_우울 [...]

기분상태_분로 [...]

기분상태_활기 [...]

기분상태_피로 [...]

기분상태_혼돈 [...]

성별=여자 [성...

결혼상태=미혼 ...

치족한련=대하

종속변수(D):

총재직의도총점 [재직의도]

블록(B) 2/ 2

이전(V)

다음(N)

블록(B) 2/ 2

병원유형=3차 병원 [병원유...]

환경-자아실현 [환경_자아...]

환경-업무량 [환경_업무량]

방법(M): 입력

통계량(S)...

도표(I)...

저장(S)...

옵션(O)...

유형(L)...

붓스트랩(B)...

선택변수(E):

규칙(U)...

케이스 레이블(C):

WLS 가중값(H):

확인

붙여넣기(P)

재설정(R)

취소

도움말

- 1단계: 일반적특성 중 단변량분석에서 유의했던 변수의 투입
- 2단계: 업무환경 하위영역 추가투입
- 3단계: 기분상태 하위영역 추가투입

선행 회귀

종속변수(D):
총재직의도총점 [재직의도]

블록(B) 3/ 3

이전(V)

다음(N)

블록(B) 3/ 3

환경-갈등 [환경_갈등]

환경-신경질 [환경_신경증]

기분상태_긴장 [긴장]

방법(M): 입력

통계량(S)...

도표(I)...

저장(S)...

옵션(O)...

유형(L)...

붓스트랩(B)...

LT00 [LT...]

병원유형_2차 ...

환경-자아실현 [...]

환경-업무량 [환...]

환경-갈등 [환경...]

환경-신경질 [환...]

총만족도평균평...

기분상태_긴장 [...]

기분상태_우울 [...]

기분상태_분로 [...]

기분상태_활기 [...]

기분상태_피로 [...]

기분상태_혼돈 [...]

성별=여자 [성...]

결혼상태=미혼 ...

최종학력=대학...

채용형태=비정...

직위=채인가호

선택변수(E):
규칙(U)...

케이스 레이블(C):

WLS 가중값(H):

확인

불여넣기(P)

재설정(R)

취소

도움말

1. (독립변수의 유의성 여부 확인 전에) 회귀모형의 적합도 및 설명력 확인
- 적합도: ANOVA table
 - 설명력: 모형요약의 R제곱 값 확인

모형 요약 ^d					
모형	R	R 제곱	수정된 R 제곱	추정값의 표준 오차	Durbin-Watson
1	.489 ^a	.239	.208	6.77121	
2	.544 ^b	.295	.235	6.65529	
3	.627 ^c	.393	.296	6.38534	1.793

a. 예측자: (상수), 병원유형=3차 병원, 채용형태=비정규직, 결혼상태=미혼, 근무병동=특수병동(ICU, OR, RR, ER)

b. 예측자: (상수), 병원유형=3차 병원, 채용형태=비정규직, 결혼상태=미혼, 근무병동=특수병동(ICU, OR, RR, ER), 환경-업무량, 환경-신경질, 환경-자아실현, 환경-갈등

c. 예측자: (상수), 병원유형=3차 병원, 채용형태=비정규직, 결혼상태=미혼, 근무병동=특수병동(ICU, OR, RR, ER), 환경-업무량, 환경-신경질, 환경-자아실현, 환경-갈등, 기분상태_활기, 기분상태_혼돈, 기분상태_우울, 기분상태_분로, 기분상태_피로, 기분상태_긴장

d. 종속변수: 총재직의도총점

1. (독립변수의 유의성 여부 확인 전에) 회귀모형의 적합도 및 설명력 확인
- 적합도: ANOVA table
 - 설명력: 모형요약의 R제곱 값 확인

모형 요약 ^d					
모형	R	R 제곱	수정된 R 제곱	추정값의 표준 오차	Durbin-Watson
1	.489 ^a	.239	.208	6.77121	
2	.544 ^b	.295	.235	6.65529	
3	.627 ^c	.393	.296	6.38534	1.793

a. 예측자: (상수), 병원유형=3차 병원, 채용형태=비정규직, 결혼상태=미혼, 근무병동=특수병동(ICU, OR, RR, ER)

b. 예측자: (상수), 병원유형=3차 병원, 채용형태=비정규직, 결혼상태=미혼, 근무병동=특수병동(ICU, OR, RR, ER), 환경-업무량, 환경-신경질, 환경-자아실현, 환경-갈등

c. 예측자: (상수), 병원유형=3차 병원, 채용형태=비정규직, 결혼상태=미혼, 근무병동=특수병동(ICU, OR, RR, ER), 환경-업무량, 환경-신경질, 환경-자아실현, 환경-갈등, 기분상태_활기, 기분상태_혼돈, 기분상태_우울, 기분상태_분로, 기분상태_피로, 기분상태_긴장

d. 종속변수: 총재직의도총점

1. (독립변수의 유의성 여부 확인 전에) 회귀모형의 적합도 및 설명력 확인 (cont'd)
- 적합도: ANOVA table
 - 설명력: 모형요약의 R제곱 값 확인

ANOVA ^a						
모형		제곱합	자유도	평균제곱	F	유의확률
1	회귀	1398.780	4	349.695	7.627	<.001 ^b
	잔차	4447.387	97	45.849		
	전체	5846.167	101			
2	회귀	1726.927	8	215.866	4.874	<.001 ^c
	잔차	4119.240	93	44.293		
	전체	5846.167	101			
3	회귀	2298.953	14	164.211	4.027	<.001 ^d
	잔차	3547.214	87	40.773		
	전체	5846.167	101			

- a. 종속변수: 총재직의도총점
- b. 예측자: (상수), 병원유형=3차 병원, 채용형태=비정규직, 결혼상태=미혼, 근무병동=특수병동(ICU, OR, RR, ER)
- c. 예측자: (상수), 병원유형=3차 병원, 채용형태=비정규직, 결혼상태=미혼, 근무병동=특수병동(ICU, OR, RR, ER), 환경-업무량, 환경-신경질, 환경-자아실현, 환경-갈등
- d. 예측자: (상수), 병원유형=3차 병원, 채용형태=비정규직, 결혼상태=미혼, 근무병동=특수병동(ICU, OR, RR, ER), 환경-업무량, 환경-신경질, 환경-자아실현, 환경-갈등, 기분상태_활기, 기분상태_혼돈, 기분상태_우울, 기분상태_분로, 기분상태_피로, 기분상태_긴장

DataSet 검토

Dummy Variable 만들기

단순회귀분석

다중회귀분석

위계적 회귀분석

기본 개념과 연구가설

SPSS로 분석하기

SPSS 출력결과표 해석

논문결과표 작성 및 기술

모형		비표준화 계수		표준화 계수	t	유의확률	공선성 통계량	VIF
		B	표준화 오류	베타			공차	
1	(상수)	37.454	1.252		29.920	<.001		
	결혼상태=미혼	-2.537	1.416	-.162	-1.791	.076	.958	1.043
	채용형태=비정규직	-2.746	1.737	-.141	-1.581	.117	.983	1.017
	근무병동=특수병동(ICU, OR, RR, ER)	-3.189	1.484	-.210	-2.149	.034	.824	1.213
	병원유형=3차 병원	-3.872	1.491	-.255	-2.597	.011	.817	1.224
2	(상수)	35.101	8.229		4.266	<.001		
	결혼상태=미혼	-2.836	1.523	-.181	-1.861	.066	.800	1.249
	채용형태=비정규직	-.909	1.858	-.047	-.489	.626	.830	1.205
	근무병동=특수병동(ICU, OR, RR, ER)	-2.815	1.581	-.185	-1.780	.078	.702	1.425
	병원유형=3차 병원	-3.517	1.492	-.231	-2.357	.021	.787	1.270
	환경-자아실현	2.829	1.476	.185	1.916	.058	.811	1.233
	환경-업무량	-.317	2.358	-.013	-.134	.893	.785	1.274
	환경-갈등	-1.581	1.530	-.120	-1.034	.304	.565	1.770
3	(상수)	28.575	8.346		3.424	<.001		
	결혼상태=미혼	-2.590	1.503	-.165	-1.723	.088	.757	1.321
	채용형태=비정규직	.826	2.044	.043	.404	.687	.631	1.585
	근무병동=특수병동(ICU, OR, RR, ER)	-2.150	1.596	-.141	-1.347	.181	.634	1.577
	병원유형=3차 병원	-2.340	1.708	-.154	-1.370	.174	.554	1.806
	환경-자아실현	1.237	1.527	.081	.810	.420	.698	1.432
	환경-업무량	.770	2.301	.032	.335	.739	.759	1.318
	환경-갈등	-1.675	1.553	-.127	-1.079	.284	.505	1.981
	환경-신경질	1.030	1.165	.089	.885	.379	.690	1.450
	기분상태_긴장	-.043	.342	-.021	-.127	.899	.258	3.876
	기분상태_우울	-.258	.228	-.153	-1.133	.260	.383	2.609
	기분상태_분로	-.169	.325	-.077	-.519	.605	.315	3.176
	기분상태_활기	.689	.245	.305	2.813	.006	.593	1.685
	기분상태_피로	.088	.250	.053	.351	.727	.306	3.268
	기분상태_혼돈	-.110	.396	-.049	-.278	.782	.227	4.409

a. 종속변수: 총재직의도총점

[표 3] 업무환경_자아실현이 재직의도에 미치는 영향

(N=102)

독립변수	MODEL 1						MODEL 2						MODEL 3					
	B	SE	β	t	p	VIF	B	SE	β	t	p	VIF	B	SE	β	t	p	VIF
(상수)	37.454	1.252		29.920	0.000		35.101	8.229		4.266	0.000		28.575	8.346		3.424	0.001	
결혼상태_미혼	-2.537	1.416	-0.162	-1.791	0.076	1.043	-2.836	1.523	-0.181	-1.861	0.066	1.249	-2.590	1.503	-0.165	-1.723	0.088	1.321
채용형태_비정규직	-2.746	1.737	-0.141	-1.581	0.117	1.017	-0.909	1.858	-0.047	-0.489	0.626	1.205	0.826	2.044	0.043	0.404	0.687	1.585
근무병동_특수병동	-3.189	1.484	-0.210	-2.149	0.034	1.213	-2.815	1.581	-0.185	-1.780	0.078	1.425	-2.150	1.596	-0.141	-1.347	0.181	1.577
병원유형_3차병원	-3.872	1.491	-0.255	-2.597	0.011	1.224	-3.517	1.492	-0.231	-2.357	0.021	1.270	-2.340	1.708	-0.154	-1.370	0.174	1.806
업무환경_자아실현							2.829	1.476	0.185	1.916	0.058	1.233	1.237	1.527	0.081	0.810	0.420	1.432
업무환경_업무량							-0.317	2.358	-0.013	-0.134	0.893	1.274	0.770	2.301	0.032	0.335	0.739	1.318
업무환경_갈등							-1.581	1.530	-0.120	-1.034	0.304	1.770	-1.675	1.553	-0.127	-1.079	0.284	1.981
업무환경_신경증							-0.438	1.095	-0.038	-0.400	0.690	1.180	1.030	1.165	0.089	0.885	0.379	1.450
기분상태_긴장													-0.043	0.342	-0.021	-0.127	0.899	3.876
기분상태_우울													-0.258	0.228	-0.153	-1.133	0.260	2.609
기분상태_분노													-0.169	0.325	-0.077	-0.519	0.605	3.176
기분상태_활기													0.689	0.245	0.305	2.813	0.006	1.685
기분상태_피로													0.088	0.250	0.053	0.351	0.727	3.268
기분상태_혼돈													-0.110	0.396	-0.049	-0.278	0.782	4.409
$F=7.627 (p<.001), R^2=.239, adj.R^2=.208$							$F=4.874 (p<.001), R^2=.295, adj.R^2=.235$						$F=4.027 (p<.001), R^2=.393, adj.R^2=.296, D-W=1.793$					

DataSet 검토

Dummy Variable 만들기

단순회귀분석

다중회귀분석

위계적 회귀분석

기본 개념과 연구가설

SPSS로 분석하기

SPSS 출력결과표 해석

논문결과표 작성 및 기술

간호사의 일반적특성과 간호사가 인지한 업무환경 및 간호사의 기분상태가 재직의도에 미치는 영향을 검증하기 위해 위계적 회귀분석을 실시한 결과가 [표 3]에 제시되어 있다. 회귀모형은 1단계($F=4.181, p<.001$), 2단계($F=4.181, p<.001$), 3단계($F=4.181, p<.001$)에서 모두 통계적으로 유의하였으며, 회귀모형의 설명력은 1단계에서 약 20.8%, 2단계에서 약 23.5%, 3단계에서 약 29.6%로 나타났다. Durbin-Watson 통계량은 1.793으로 2에 근사한 값을 보여 잔차의 독립성 가정에 문제가 없는 것으로 평가되었고, 분산팽창지수(Variance Inflation Factor, VIF)도 모두 10미만으로 나타나 다중공선성 문제는 없는 것으로 판단되었다.

(다음 슬라이드에 계속)

[표 3] 업무환경_자아실현이 재직의도에 미치는 영향

(N=102)

독립변수	MODEL 1						MODEL 2						MODEL 3					
	B	SE	β	t	p	VIF	B	SE	β	t	p	VIF	B	SE	β	t	p	VIF
(상수)	37.454	1.252		29.920	0.000		35.101	8.229		4.266	0.000		28.575	8.346		3.424	0.001	
결혼상태_미혼	-2.537	1.416	-0.162	-1.791	0.076	1.043	-2.836	1.523	-0.181	-1.861	0.066	1.249	-2.590	1.503	-0.165	-1.723	0.088	1.321
채용형태_비정규직	-2.746	1.737	-0.141	-1.581	0.117	1.017	-0.909	1.858	-0.047	-0.489	0.626	1.205	0.826	2.044	0.043	0.404	0.687	1.585
근무병동_특수병동	-3.189	1.484	-0.210	-2.149	0.034	1.213	-2.815	1.581	-0.185	-1.780	0.078	1.425	-2.150	1.596	-0.141	-1.347	0.181	1.577
병원유형_3차병원	-3.872	1.491	-0.255	-2.597	0.011	1.224	-3.517	1.492	-0.231	-2.357	0.021	1.270	-2.340	1.708	-0.154	-1.370	0.174	1.806
업무환경_자아실현							2.829	1.476	0.185	1.916	0.058	1.233	1.237	1.527	0.081	0.810	0.420	1.432
업무환경_업무량							-0.317	2.358	-0.013	-0.134	0.893	1.274	0.770	2.301	0.032	0.335	0.739	1.318
업무환경_갈등							-1.581	1.530	-0.120	-1.034	0.304	1.770	-1.675	1.553	-0.127	-1.079	0.284	1.981
업무환경_신경증							-0.438	1.095	-0.038	-0.400	0.690	1.180	1.030	1.165	0.089	0.885	0.379	1.450
기분상태_긴장													-0.043	0.342	-0.021	-0.127	0.899	3.876
기분상태_우울													-0.258	0.228	-0.153	-1.133	0.260	2.609
기분상태_분노													-0.169	0.325	-0.077	-0.519	0.605	3.176
기분상태_활기													0.689	0.245	0.305	2.813	0.006	1.685
기분상태_피로													0.088	0.250	0.053	0.351	0.727	3.268
기분상태_혼돈													-0.110	0.396	-0.049	-0.278	0.782	4.409
F=7.627 (p<.001), R ² =.239, adj.R ² =.208							F=4.874 (p<.001), R ² =.295, adj.R ² =.235						F=4.027 (p<.001), R ² =.393, adj.R ² =.296, D-W=1.793					

DataSet 검토

Dummy Variable 만들기

단순회귀분석

다중회귀분석

위계적 회귀분석

기본 개념과 연구가설

SPSS로 분석하기

SPSS 출력결과표 해석

논문결과표 작성 및 기술

(앞 슬라이드에서 계속)

회귀계수의 유의성 검증결과, 1단계에서는 근무병동이 특수병동인 경우 ($\beta=-.210, p=.034$) 와 병원유형이 3차 병원인 경우 ($\beta=-.255, p=.011$) 재직의도에 부(-)적으로 유의하게 나타났다. 즉, 일반병동에 근무하는 간호사보다 특수병동에 근무하는 간호사와 2차병원에 근무하는 간호사보다 3차병원에 근무하는 간호사의 재직의도가 유의하게 낮은 것으로 평가되었다.

2단계에는, 통제변수 중 병원유형이 3차 병원인 경우 ($\beta=-.231, p=.021$) 재직의도에 부(-)적으로 유의하게 나타났고, 업무환경의 하위영역은 모두 유의한 영향을 미치지 못하는 것으로 나타났다. 즉, 일반병동에 근무하는 간호사보다 특수병동에 근무하는 간호사의 재직의도가 유의하게 낮은 것으로 평가되었다.

3단계에서는, 통제변수와 업무환경의 하위영역은 모두 유의하지 않았고, 기분상태 하위영역 중 활기($\beta=-2.813, p=.006$)가 재직의도에 유의한 정(+)의 영향을 미치는 것으로 나타났다. 즉, 간호사의 기분상태 중 활기가 높아질 수록 간호사의 재직의도도 높아지는 것으로 평가되었다.

감사합니다

Joyhippo@hanyang.ac.kr

04

로지스틱 회귀분석

문설화 교수

한양대학교 간호과학연구소



로지스틱 회귀분석 (Logistic regression analysis)

2022.06.30.

문설화(RN, GNP, PhD)

(e-mail: seora@hanyang.ac.kr)

Contents

1. Logistic regression 소개

2. DATA 이용한 SPSS실습

- Data 확인
- Univariable analysis
- Multivariable analysis

Table 2. Logistic Regression Models for Suicidal Ideation across Life Cycle

(N=7,000)

Variables	Categories	Young adults (n=2,549)		Middle-aged adults (n=3,422)		Older adults (n=1,029)	
		Univariable	Multivariable	Univariable	Multivariable	Univariable	Multivariable
		Crude OR (95% CI)	Adjusted OR (95% CI)	Crude OR (95% CI)	Adjusted OR (95% CI)	Crude OR (95% CI)	Adjusted OR (95% CI)
Gender	Male (ref. Female)	1.00 (0.69~1.45)		0.97 (0.74~1.27)		1.14 (0.76~1.70)	
Education level completed	≤ High school (ref. ≥ College)	2.33 (1.58~3.43)***	1.62 (1.05~2.49)*	1.85 (1.34~2.57)***	1.58 (1.12~2.23)**	1.00 (0.35~2.86)	
Presence of spouse	No (ref. Yes)	1.33 (0.90~1.95)		3.67 (2.58~5.20)***	2.55 (1.72~3.78)***	2.60 (1.74~3.89)***	2.16 (1.37~3.39)**
Monthly household income (won)	< 4 million (ref. ≥ 4 million)	1.76 (1.20~2.59)*		1.73 (1.31~2.29)***		1.19 (0.53~2.65)	
Self-esteem		1.36 (1.28~1.44)***	1.19 (1.12~1.27)***	1.27 (1.22~1.32)***	1.13 (1.08~1.18)***	1.37 (1.28~1.47)***	1.22 (1.13~1.32)***
Stress level	High (ref. Low)	4.71 (3.05~7.27)***	3.26 (2.07~5.15)***	3.50 (2.64~4.65)***	2.46 (1.81~3.33)***	4.52 (2.99~6.82)***	2.21 (1.38~3.54)**

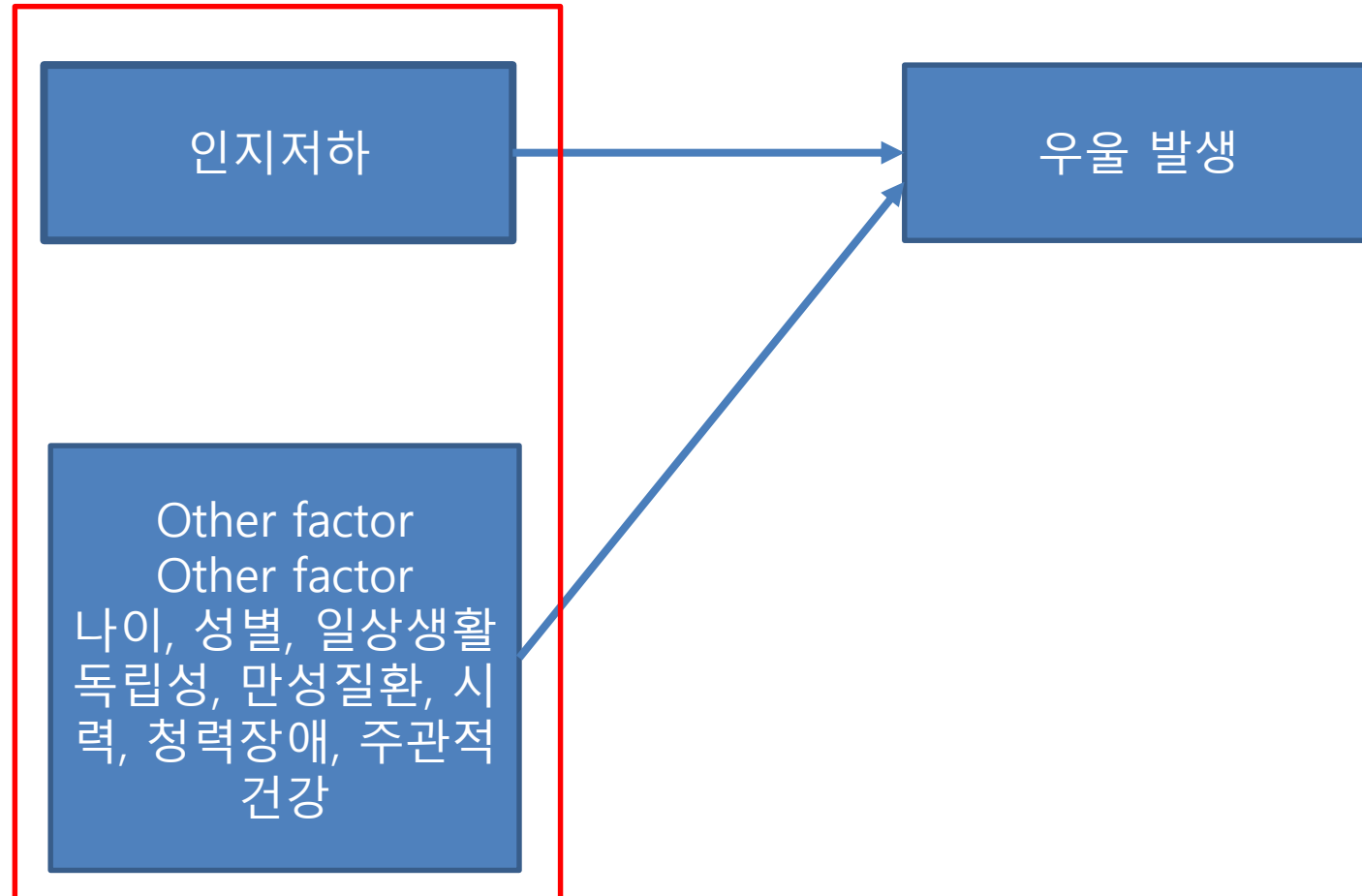
- Self esteem이 낮을 수록 자살의도는 높아진다...?
- 스트레스수준이 낮은 사람보다 높은 사람이 자살의도가 4.71배 높다??
- 95% CI??
- OR??
- Univariable vs Multivariable??

Research Question

- 인지저하가 우울의 예측요인 일까?
- 그 외의 우울예측요인은 무엇이 있을까?

독립변수, 설명변수

종속변수, 반응변수



Logistic regression model

회귀식: $\log \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m$

- p = 사건이 발생할 확률(우울이 발생할 확률)
- Odds(오즈) = $p/(1-p) \rightarrow$ 우울 발생확률이 그렇지 않을 확률에 비해 몇배 더 높은지 의미

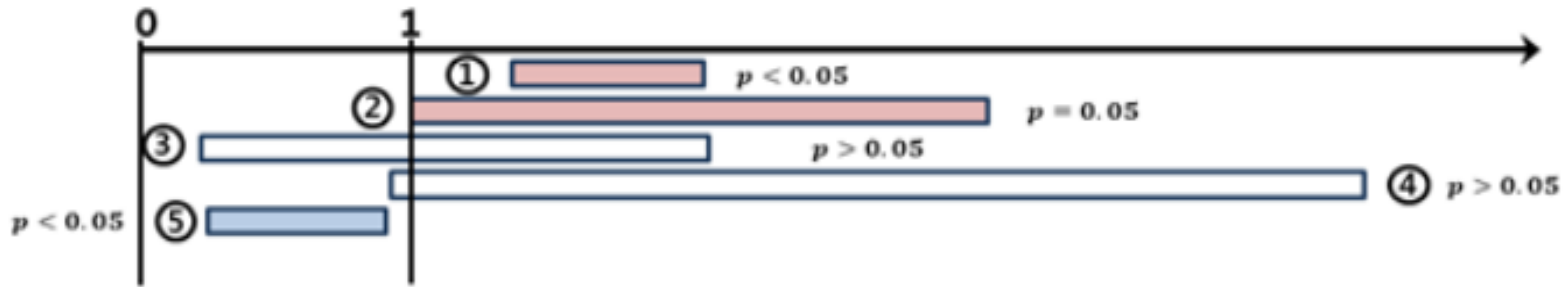
- Cog_impaired = 1 / normal = 0
- normal 이면,
 $\ln(\text{odds}(\text{normal})) = \beta_0 + \beta_1 \times 0 = \beta_0$
- Cog_impaired 이며
 $\ln(\text{odds}(\text{Cog_impaired})) = \beta_0 + \beta_1 \times 1 = \beta_0 + \beta_1$
 - $\rightarrow \ln(\text{odds}(\text{Cog_impaired})) - \ln(\text{odds}(\text{normal})) = \beta_1$
 - $\rightarrow \ln(\text{odds}(\text{Cog_impaired}) / (\text{odds}(\text{normal}))) = \ln(\text{OR}(\text{Cog_impaired})) = \beta_1$
 - $\rightarrow \text{OR}(\text{Cog_impaired}) = e^{\beta_1} = \text{exp}(\beta_1)$

$\text{exp}(\beta_1)$: 인지저하가 없는 사람에 비해 인지저하가 있는 사람에서 우울 발생 비율이 $\text{exp}(\beta_1)$ 배 이다.

교차비(odd ratio)의 신뢰구간

OR=교차비=승산비

교차비(Odds ratio)의 95% 신뢰구간



- 교차비: 질병과 개별 위험인자 사이의 연관성 정도 의미
- 1: **질병과 개별 위험 인자가 연관성이 없음**을 의미
- ①⑤ : 교차비의 모수가 해당구간에 있을 가능성이 95% → 실제 교차비가 1이어서 질병과 위험인자 사이에 연관이 전혀 없을 가능성은 5% 미만($p < .05$)이라고 해석.
- ③④: 교차비의 모수의 95% CI가 1을 포함하고 있으며 질병과 위험인자의 연관성이 없을 가능성 (교차비=1)이 신뢰구간 내에 포함되고 있으므로 질병의 위험이 유의하게 증가한 다고 말할 수 없음($p > .05$).

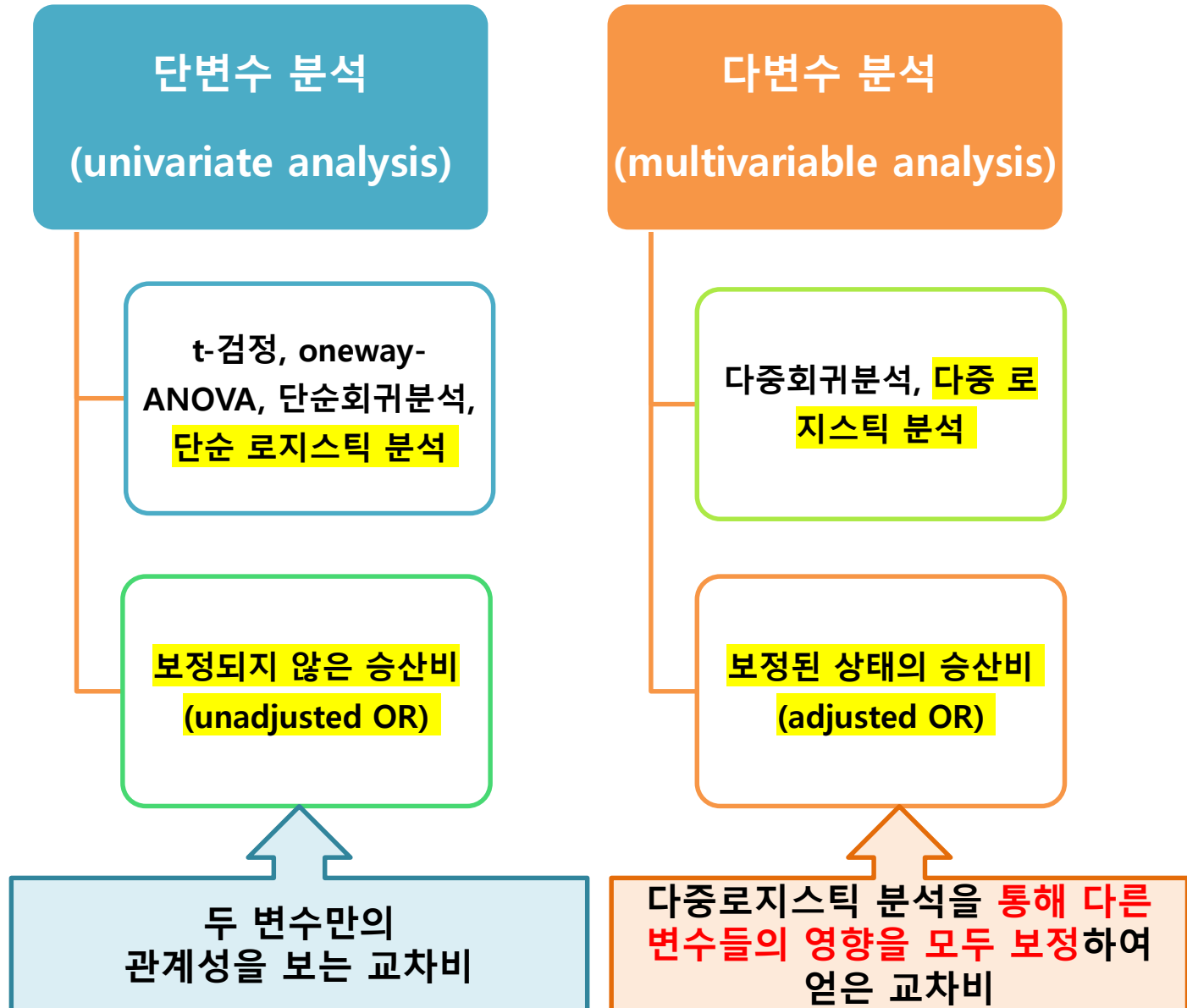
Odds ratio의 의미

β_1	$\exp(\beta_1)=OR$	Meaning
0	1 Odd(cog_impaired)=odd(normal)	인지저하 여부가 우울발생에 영향을 미치지 않음
>0	>1 Odd(cog_impaired)>odd(normal) (ex. OR=2)	정상인 그룹 대비 인지저하 그룹에서 우울이 발생할 가능성이 $\exp(\beta_1)=2$ 배 이다
<0	<1 Odd(cog_impaired)<odd(normal) (ex. OR=0.5)	인지저하그룹대비 정상인 그룹에서 우울이 발생할 가능성이 $\exp(\beta_1)=2$ 배 이다

** 독립변수가 연속형 변수일때 (ex. Age)

$\exp(\beta_1)$: age가 1단위 증가할 때 우울이 발생할 가능성이 $\exp(\beta_1)$ 배 증가한다.

Univariate vs Multivariate analysis



Univariate vs Multivariate analysis

단변수 분석

(univariate analysis)

↑
유의한 관련 변수 확인!!!!

다변수 분석

(multivariable analysis)

↑
단변수 분석에서 유의한 변수만 투입
→ multicollinearity(다중공선성) 고려하여 변수 선택
→ 최종 model에는 단변수 분석/다중공선성 고려한 변수만 포함!!

Multiple linear regression vs Multiple logistic regression

	다중회귀분석(MR)	다중 로지스틱회귀분석(MLR)
종속변수	연속형	범주형
독립변수	연속형, 범주형 모두 가능	
Model fit(회귀 모형 평가)	Adj R ²	Nagelkerke R ² Hosmer-Lemeshow test
얼고자 하는 값	회귀계수	승산비(exp)의 95% CI(신뢰구간)

Contents

1. Logistic regression 소개
2. DATA 이용한 SPSS실습
 - Data 확인
 - Univariable analysis
 - Multivariable analysis

Variable_coding book

변수 구분	변수명	변수 측 도	설명	범위	새로운 변수명	코딩 변경	
						0	1
독립 변수	sex	비연속	성별	0=남자, 1=여자		그대로	
	age	연속	만나이	만나이(85세 이상~)		그대로	
	IADL_2g	비연속	일상생활독립성(IADL)	10개 문항(몸단장~교통수단이용) →모두 독립인경우만 독립 0=모든항목 독립, 1=1항목 이상 의존		그대로	
	ch_dz_total	연속	만성질환 수		0~9	ch_dz_2g	2개 이하 3개 이상
	C1_3	비연속	청력_일상생활불편함	0=불편하지 않음, 1=불편함	0~11		그대로
	Cognition_2g	비연속	인지저하 유무	0=인지저하 없음 1=인지저하 있음			그대로
종속 변수	GDS15_sum	연속	우울(GSD-15)	지난 일주일동안의 생활에 대해: 15문항 (예=1, 아니오=0) 문항당 1점 총점	0~12	gds_15_2g	0-7점 8점이상

Tip: 변수를 범주화 할 때(recoding)는 문헌에 근거해서 해야함!!!

코딩 변경: 만성질환 수

변수명	변경 범주
ch_dz_2g	0=2개 이하, 1=3개 이상

변수명	코딩 변경	
ch_dz_2g	0	1
	2개 이하	3개 이상

새로운 변수로 코딩 변경

- 대상자ID [PID]
- 만연령 [age]
- 성별 [sex]
- living_alone [livin...]
- edu_2g [edu_2g]
- IADL_2g [IADL_2g]
- 시력_일상생활의 ...
- 청력_일상생활의 ...
- 평소의 건강상태 [...]
- GDS_15문항 합...
- ch_dz_2g [ch_dz...]
- perceived_health...
- cognition_2g [cog...]
- GDS_15_2g [GD...

숫자 변수(V) -> 출력변수:

ch_dz_total -> ch_dz_2g

출력변수

이름(N):

ch_dz_2g

설명(L):

만성질환_2g

바꾸기(B)

새로운 변수로 코딩 변경: 기존값 및 새로운 값

기존값

☐ 값(V):

☐ 시스템-결측값(S)

☐ 시스템 또는 사용자 결측값(U)

☐ 범위(N):

에서(T)

☐ 최저값에서 다음 값까지 범위(G):

☒ 다음 값에서 최고값까지 범위(E):

☐ 기타 모든 값(O)

새로운 값

☒ 기존값(A):

☐ 시스템-결측값(Y)

☐ 기존값 복사(P)

기존값 -> 새로운 값(D):

Lowest thru 2 -> 0

3 thru Highest -> 1

추가(A)

바꾸기(C)

제거(R)

☐ 출력변수가 문자열임(B)

너비(W):

8

☐ 숫자형 문자를 숫자로 변환('5' -> 5)(M)

계속

취소

도움말

코딩 변경: 우울

변수명	코딩 변경	
gds_15_2g	0	1
	0-7점	8점이상

숫자	8	2	living_a
숫자	8	2	edu_2g
숫자	8	2	IADL_2
숫자	12	0	만성질환
숫자	12	0	시력_을

새로운 변수로 코딩변경

숫자 변수(V) -> 출력변수:

GDS15_sum -> GDS15_2g

대상자ID [PID]

만연령 [age]

성별 [sex]

출력변수

이름(N): GDS15_2g

설명(L): GDS15_2g

바꾸기(B)

새로운 변수로 코딩변경: 기존값 및 새로운 값

기존값

☒ 값(V):

☐ 시스템-결측값(S)

☐ 시스템 또는 사용자 결측값(U)

☐ 범위(N):

에서(I)

☐ 최저값에서 다음 값까지 범위(G):

☒ 다음 값에서 최고값까지 범위(E):

☐ 기타 모든 값(O)

새로운 값

☒ 기존값(A):

☐ 시스템-결측값(Y)

☐ 기존값 복사(P)

기존값 -> 새로운 값(D):

Lowest thru 7 -> 0

8 thru Highest -> 1

추가(A)

바꾸기(C)

제거(R)

☐ 출력변수가 문자열임(B) 너비(W): 8

☐ 숫자형 문자를 숫자로 변환('5' -> 5)(M)

계속

취소

도움말

생성 변수에 대한 변수값 설명입력

입(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)



이름	유형	너비	소수점이...	설명	값	결측값	열	맞춤	측도	역할
PID	숫자	8	0	대상자ID	없음	없음	10	오른쪽	척도(S)	입력
age	숫자	8	0	만연령	{999, 비해당}...	999	10	오른쪽	척도(S)	입력
sex	숫자	8	0	성별	{0, 남자}...	9999	10	오른쪽	명목(N)	입력
living_alone	숫자	8	2	living_alone	{00, 동거}...	없음	14	오른쪽	명목(N)	입력
edu_2g	숫자	8	2	edu_2g	{00, 7년이상(중등이상)}...	없음	10	오른쪽	명목(N)	입력
IADL_2g	숫자	8	2	IADL_2g	{00, 독립}...	없음	10	오른쪽	명목(N)	입력
ch_dz_total	숫자	12	0	만성질환 수	없음	없음	12	오른쪽	명목(N)	입력
C1_2	숫자	12	0	시력_일상생...	{0, 불편				명목(N)	입력
C2_2	숫자	12	0	청력_일상생...	{0, 불편				명목(N)	입력
perceived_h...	숫자	12	0	주관적 건강...	{1, 매우				명목(N)	입력
GDS15_sum	숫자	8	2	GDS_15문항 ...	없음				척도(S)	입력
ch_dz_2g	숫자	8	2	만성질환_2g	{00, 2개				명목(N)	입력
perceived_h...	숫자	8	2	주관적 건강_2g	{00, go				명목(N)	입력
cognitive_2g	숫자	8	2	cognitive_2g	언어 ...				명목(N)	입력

우울

변수값 설명

기준값(A):

맞춤법(S)...

설명(L):

.00 = "정상(0-7)"
1.00 = "우울(8이상)"

추가(A)

바꾸기(C)

제거(R)

확인

취소

도움말

변수값 설명

만성질환

변수값 설명

기준값(A):

맞춤법(S)...

설명(L):

.00 = "2개이하"
1.00 = "3개이상"

추가(A)

바꾸기(C)

제거(R)

확인

취소

도움말

각 변수의 빈도 확인

만연령

	빈도	퍼센트	유효 퍼센트	누적퍼센트
유효 85	155	25.4	25.4	25.4
86	115	18.9	18.9	44.3
87	92	15.1	15.1	59.3
88	70	11.5	11.5	70.8
89	51	8.4	8.4	79.2
90	38	6.2	6.2	85.4
91	27	4.4	4.4	89.8
92	24	3.9	3.9	93.8
93	13	2.1	2.1	95.9
94	9	1.5	1.5	97.4
95	8	1.3	1.3	98.7
96	4	.7	.7	99.3
97	2	.3	.3	99.7
98	1	.2	.2	99.8
99	1	.2	.2	100.0
합계	610	100.0	100.0	

성별

	빈도	퍼센트	유효 퍼센트	누적퍼센트
유효 남자	168	27.5	27.5	27.5
여자	442	72.5	72.5	100.0
합계	610	100.0	100.0	

IADL_2g

	빈도	퍼센트	유효 퍼센트	누적퍼센트
유효 독립	399	65.4	65.4	65.4
의존	211	34.6	34.6	100.0
합계	610	100.0	100.0	

만성질환_2g

	빈도	퍼센트	유효 퍼센트	누적퍼센트
유효 2개이하	372	61.0	61.0	61.0
3개이상	238	39.0	39.0	100.0
합계	610	100.0	100.0	

정력_일상생활의 불편함

	빈도	퍼센트	유효 퍼센트	누적퍼센트
유효 불편하지 않음	299	49.0	49.0	49.0
불편함	311	51.0	51.0	100.0
합계	610	100.0	100.0	

인지_2g

	빈도	퍼센트	유효 퍼센트	누적퍼센트
유효 인지저하 없음	280	45.9	45.9	45.9
인지저하 있음	330	54.1	54.1	100.0
합계	610	100.0	100.0	

우울_2g

	빈도	퍼센트	유효 퍼센트	누적퍼센트
유효 정상(0-7)	552	90.5	90.5	90.5
우울(8이상)	58	9.5	9.5	100.0
합계	610	100.0	100.0	

종속변수 & 각 변수의 교차표 확인

분석-기술통계-교차분석

우울_2g * 성별 교차표

			성별		전체
			남자	여자	
우울_2g	정상(0-7)	빈도	153	399	552
		우울_2g 중 %	27.7%	72.3%	100.0%
		성별 중 %	91.1%	90.3%	90.5%
	우울(8이상)	빈도	15	43	58
		우울_2g 중 %	25.9%	74.1%	100.0%
		성별 중 %	8.9%	9.7%	9.5%
전체	빈도	168	442	610	
		우울_2g 중 %	27.5%	72.5%	100.0%
		성별 중 %	100.0%	100.0%	100.0%

카이제곱 검정

	값	자유도	점근 유의확률 (양측검정)	정확한 유의확률 (양측검정)	정확한 유의확률 (단측검정)
Pearson 카이제곱	.091 ^a	1	.764	.878	.449
연속수정 ^b	.021	1	.884		
우도비	.092	1	.762		
Fisher의 정확한 검정					
선형 대 선형결합	.090	1	.764		
유효 케이스 수	610				

a. 0 셀 (0.0%)은(는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 15.97입니다.

b. 2x2 표에 대해서만 계산됨

우울_2g * IADL_2g 교차표

			IADL_2g		전체
			독립	의존	
우울_2g	정상(0-7)	빈도	375	177	552
		우울_2g 중 %	67.9%	32.1%	100.0%
		IADL_2g 중 %	94.0%	83.9%	90.5%
	우울(8이상)	빈도	24	34	58
		우울_2g 중 %	41.4%	58.6%	100.0%
		IADL_2g 중 %	6.0%	16.1%	9.5%
전체	빈도	399	211	610	
	우울_2g 중 %	65.4%	34.6%	100.0%	
	IADL_2g 중 %	100.0%	100.0%	100.0%	

카이제곱 검정

	값	자유도	점근 유의확률 (양측검정)	정확한 유의확률 (양측검정)	정확한 유의확률 (단측검정)
Pearson 카이제곱	16.359 ^a	1	.000	.000	.000
연속수정 ^b	15.206	1	.000		
우도비	15.467	1	.000		
Fisher의 정확한 검정					
선형 대 선형결합	16.332	1	.000		
유효 케이스 수	610				

a. 0 셀 (0.0%)은(는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 20.06입니다.

b. 2x2 표에 대해서만 계산됨

종속변수 & 각 변수의 교차표 확인

분석-기술통계-교차분석

우울_2g * 만성질환_2g 교차표

			만성질환_2g		전체
			2개이하	3개이상	
우울_2g	정상(0-7)	빈도	347	205	552
		우울_2g 중 %	62.9%	37.1%	100.0%
		만성질환_2g 중 %	93.3%	86.1%	90.5%
	우울(8이상)	빈도	25	33	58
		우울_2g 중 %	43.1%	56.9%	100.0%
		만성질환_2g 중 %	6.7%	13.9%	9.5%
전체	빈도	372	238	610	
	우울_2g 중 %	61.0%	39.0%	100.0%	
	만성질환_2g 중 %	100.0%	100.0%	100.0%	

우울_2g * 청력_일상생활의 불편함 교차표

			청력_일상생활의 불편함		전체	
			불편하지 않음	불편함		
우울_2g	정상(0-7)	빈도	284	268	552	
		우울_2g 중 %	51.4%	48.6%	100.0%	
		청력_일상생활의 불편함 중 %	95.0%	86.2%	90.5%	
	우울(8이상)	빈도	15	43	58	
		우울_2g 중 %	25.9%	74.1%	100.0%	
		청력_일상생활의 불편함 중 %	5.0%	13.8%	9.5%	
전체			299	311	610	
			우울_2g 중 %	49.0%	51.0%	100.0%
			청력_일상생활의 불편함 중 %	100.0%	100.0%	100.0%

카이제곱 검정

	값	자유도	점근 유의확률 (양측검정)	정확한 유의확률 (양측검정)	정확한 유의확률 (단측검정)
Pearson 카이제곱	8.612 ^a	1	.003	.004	.003
연속수정 ^b	7.802	1	.005		
우도비	8.372	1	.004		
Fisher의 정확한 검정	8.598	1	.003		
선택 대 선택결합					
유효 케이스 수	610				

a. 0 셀 (0.0%)은(는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 22.63입니다.

b. 2x2 표에 대해서만 계산됨

카이제곱 검정

	값	자유도	점근 유의확률 (양측검정)	정확한 유의확률 (양측검정)	정확한 유의확률 (단측검정)
Pearson 카이제곱	13.750 ^a	1	.000	.000	.000
연속수정 ^b	12.745	1	.000		
우도비	14.326	1	.000		
Fisher의 정확한 검정	13.728	1	.000		
선택 대 선택결합					
유효 케이스 수	610				

a. 0 셀 (0.0%)은(는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 28.43입니다.

b. 2x2 표에 대해서만 계산됨

종속변수 & 각 변수의 교차표 확인

분석-기술통계-교차분석

우울_2g * 인지_2g 교차표

			인지_2g		전체
			인지저하 없음	인지저하 있음	
우울_2g	정상(0-7)	빈도	266	286	552
		우울_2g 중 %	48.2%	51.8%	100.0%
		인지_2g 중 %	95.0%	86.7%	90.5%
	우울(8이상)	빈도	14	44	58
		우울_2g 중 %	24.1%	75.9%	100.0%
		인지_2g 중 %	5.0%	13.3%	9.5%
전체	빈도	280	330	610	
	우울_2g 중 %	45.9%	54.1%	100.0%	
	인지_2g 중 %	100.0%	100.0%	100.0%	

카이제곱 검정

	값	자유도	점근 유의확률 (양측검정)	정확한 유의확률 (양측검정)	정확한 유의확률 (단측검정)
Pearson 카이제곱	12.226 ^a	1	.000	.000	.000
연속수정 ^b	11.276	1	.001		
우도비	12.918	1	.000		
Fisher의 정확한 검정					
선형 대 선형결합	12.206	1	.000		
유효 케이스 수	610				

a. 0 셀 (0.0%)은(는) 5보다 작은 기대 빈도를 가지는 셀입니다. 최소 기대빈도는 26.62입니다.

b. 2x2 표에 대해서만 계산됨

Contents

1. Logistic regression 소개
2. DATA 이용한 SPSS실습
 - Data 확인
 - Univariable analysis
 - Multivariable analysis

Univariable analysis: 단변량(수) 분석

분석-회귀분석-이분로지스틱(종속변수: 우울)

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) **분석(A)** 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 창(W) 도움말(H)

보고서(P) > 기술통계량(E) > 표 > 평균 비교(M) > 일반선형모형(G) > 일반화 선형 모형(Z) > 혼합 모형(X) > 상관분석(C) > **회귀분석(R)** > 자동 선형 모형화... > 선형(L)... > 곡선추정(C)... > 일부 최소제곱(S)... > **이분형 로지스틱(G)...** > 다항 로지스틱(M)... > 순서(D)... > 프로빗(P)... > 비선형(N)... > 가중추정(W)... > 2-단계 최소제곱(2)... > 최적화 척도법(CATREG)...

	이름	유형	나이	설명	값	결측값	예	맞춤	척도	역할
1	PID	숫자	8		없음	없음	10	옳은쪽	척도(S)	입력
2	age	숫자	8		{999, 비해당}...	999	10	옳은쪽	척도(S)	입력
3	sex	숫자	8		{0, 남자}...	9999	10	옳은쪽	명목(N)	입력
4	IADL_2g	숫자	8		{00, 독립}...	없음	10	옳은쪽	명목(N)	입력
5	ch_dz_2g	숫자	8		{00, 2개이하}...	없음	10	옳은쪽	명목(N)	입력
6	C2_2	숫자	12		{0, 불편하지 않음}...	9	12	옳은쪽	명목(N)	입력
7	cognition_2g	숫자	8			없음	14	옳은쪽	명목(N)	입력
8	GDS15_2g	숫자	8			없음	11	옳은쪽	명목(N)	입력
9	GDS15_sum	숫자	8			없음	11	옳은쪽	척도(S)	입력

데이터 보기(D) 변수 보기(V)

Univariable analysis: 단변량(수) 분석

비연속 변수

성별(sex)이 우울 예측 여부

- 종속변수: 우울
- 독립변수: 성별(sex)

로지스틱 회귀 모형

종속변수(D): 우울_2g [GDS15_2g]

공변량(C): sex

범주형(C)...

방법(M): 입력

확인 불여넣기(P) 재설정(R) 취소 도움말

로지스틱 회귀분석: 범주형 변수 정의

공변량(C):

범주형 공변량(T): sex(표시자(처음))

대비 바꾸기

대비(N): 표시자 바꾸기(B)

참조범주(R): ☐ 마지막(L) ☒ 처음(F)

계속 취소 도움말

범주형 독립변수인 경우 기준(참조)변수 지정

처음: 기준을 작은 값으로 지정(남=0,여=1→ 남성이 참조변수!!)

** 마지막: 큰 값을 참조변수로 하고 싶을 때 지정
→ 자신이 원하는 방향과 같은 계산이 나오도록 주의!!

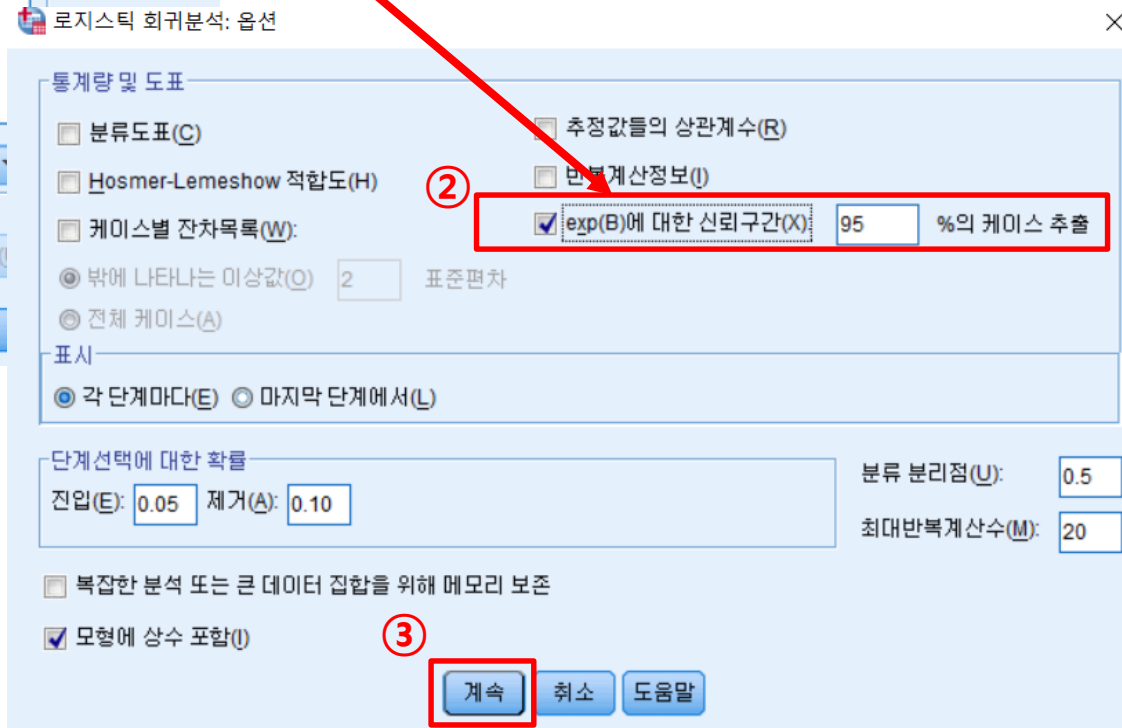
Univariable analysis: 단변량(수) 분석

비연속 변수

분석-회귀분석-이분로지스틱(종속변수: 우울)



OR값의 95% 신뢰구간 출력



** 참고: Hosmer-Lemeshow 적합도는 독립변수가 2개 이상인 경우에만 계산됨

케이스 처리 요약

가중되지 않은 케이스 ^a	N	퍼센트
선택 케이스 분석에 포함	610	100.0
결측 케이스	0	.0
합계	610	100.0
비선택 케이스	0	.0
합계	610	100.0

a. 가중값을 사용하는 경우에는 전체 케이스 수의 분류표를 참조하십시오.

방정식에 포함된 변수

	B	S.E.	Wals	자유도	유의확률	Exp(B)	EXP(B)에 대한 95% 신뢰구간	
							하한	상한
1 단계 ^a sex	.095	.315	.090	1	.764	1.099	.593	2.036
상수항	-2.322	.271	73.679	1	.000	.098		

a. 변수가 1: 단계에 진입했습니다 sex. sex.

- Sex의 유의확률이(p) .764: 성별이 우울과 통계적으로 유의하지 않으므로 related factor는 아님
- 결과 해석: 남성에 비해 여성이 우울이 발생할 가능성은 1.10배이며, 이에 대한 95% 신뢰구간(CI)은 (0.59-2.04), 유의수준(p)는 .764이다.

Univariable analysis: 단변량(수) 분석

연속 변수

성별(sex)이 우울 예측 여부

- 종속변수: 우울
- 독립변수: 나이(age)

Age는 연속변수 이므로 범주에 대한 정의를 할 필요가 없어요~!

로지스틱 회귀 모형

종속변수(D): 우울_2g [GDS15_2g]

범주형(C)... 저장(S)... 옵션(O)... 붓스트랩(T)...

블록 1 대상 1

공변량(C): age

로지스틱 회귀분석: 옵션

통계량 및 도표

- ☐ 분류도표(C)
- ☐ Hosmer-Lemeshow 적합도(H)
- ☐ 케이스별 잔차목록(W):
- ☒ exp(B)에 대한 신뢰구간(X) 95 %의 케이스 추출
- ☐ 추정값들의 상관관계수(R)
- ☐ 반복계산정보(I)

표시

- ☒ 각 단계마다(E) ☐ 마지막 단계에서(L)

단계선택에 대한 확률

진입(E): 0.05 제거(A): 0.10

분류 분리점(U): 0.5

최대반복계산수(M): 20

☐ 복잡한 분석 또는 큰 데이터 집합을 위해 메모리 보존

☒ 모형에 상수 포함(I)

계속 취소 도움말

4 확인

1 옵션(O)...

2

3

방정식에 포함된 변수

	B	S.E.	Wals	자유도	유의확률	Exp(B)	EXP(B)에 대한 95% 신뢰구간	
							하한	상한
1 단계 ^a age	.020	.050	.151	1	.698	1.020	.924	1.125
상수항	-3.966	4.413	.808	1	.369	.019		

a. 변수가 1: 단계에 진입했습니다 age. age.

- Age 변수의 유의 확률은 .698로 나이가 우울 발생에 통계적으로 유의하지 않음→나이는 우울의 related facto는 아님을 확인
- 결과 해석: age의 OR이 1.02이므로 나이가 1살 증가할수록 우울 발생할 가능상이 1.02배 증가하며, 이때 95% 신뢰구간은 (0.92-1.13)으로 유의확률(p)는 .698 이다.

Univariable analysis table

나머지 변수에 대해 통계를 실행해 보고 다음의 table을 정리해 봅시다~

Variable	OR	95% CI	<i>p</i>
Sex(성별)	1.10	0.59-2.04	.764
Age(나이)	1.02	0.92-1.13	.698
IADL_2g(독립성)			
ch_dz_total(만성질환)			
C1_3(청력불편감)			
Cognition_2g(인지장애)			

단변량 분석에서 OR
=Crude OR
=Unadjusted OR

방정식에 포함된 변수

IADL(독립성)

	B	S.E.	Wals	자유도	유의확률	Exp(B)	EXP(B)에 대한 95% 신뢰구간	
							하한	상한
1 단계 ^a IADL_2g(1)	1.099	.282	15.215	1	.000	3.001	1.728	5.214
상수항	-2.749	.211	170.443	1	.000	.064		

a. 변수가 1: 단계에 진입했습니다 IADL_2g. IADL_2g.

방정식에 포함된 변수

Ch_dz_2g
(만성질환)

	B	S.E.	Wals	자유도	유의확률	Exp(B)	EXP(B)에 대한 95% 신뢰구간	
							하한	상한
1 단계 ^a ch_dz_2g(1)	.804	.279	8.280	1	.004	2.234	1.292	3.863
상수항	-2.630	.207	161.356	1	.000	.072		

a. 변수가 1: 단계에 진입했습니다 ch_dz_2g. ch_dz_2g.

방정식에 포함된 변수

C2_2
(청력_불편감)

	B	S.E.	Wals	자유도	유의확률	Exp(B)	EXP(B)에 대한 95% 신뢰구간	
							하한	상한
1 단계 ^a C2_2(1)	1.111	.312	12.705	1	.000	3.038	1.649	5.596
상수항	-2.941	.265	123.227	1	.000	.053		

a. 변수가 1: 단계에 진입했습니다 C2_2. C2_2.

방정식에 포함된 변수

Cognition
(인지장애)

	B	S.E.	Wals	자유도	유의확률	Exp(B)	EXP(B)에 대한 95% 신뢰구간	
							하한	상한
1 단계 ^a cognition_2g(1)	1.073	.318	11.345	1	.001	2.923	1.566	5.456
상수항	-2.944	.274	115.307	1	.000	.053		

a. 변수가 1: 단계에 진입했습니다 cognition_2g. cognition_2g.



단변수(량) 분석 결과 p 값이 유의미한 변수

→ 독립성, 만성질환, 청력 불편감, 인지장애

Contents

1. Logistic regression 소개
2. DATA 이용한 SPSS실습
 - Data 확인
 - Univariable analysis
 - Multivariable analysis

Multivariable analysis: 다변량(수) 분석

Multiple logistic analysis model 실시 전 고려할 사항!!

- 다중 공선성(multicollinearity): 일부 독립변수가 다른 변수와의 상관성이 높은 경우 데이터 분석 시 부정적 영향을 미쳐 각 독립변수의 개별 효과를 파악하기 어려움. → 독립변수간의 상관성 확인 후 상관성이 높은 변수는 제외하고 분석할 필요가 있음
- 변수 선택(variable selection): 단변량(수) 분석에서 유의한 변수를 최종 모형에 포함.

Multivariable analysis: 다변량(수) 분석

다중 공선성(multicollinearity)

- VIF(Variance Inflation Factor) 확인 → 일반적으로 VIF 값이 10을 넘지 않으면 모형에 포함.
- Multiple regression 분석에서 확인 가능함
(회귀분석 > 선형 > 통계량 > 공선성진단)

선형 회귀분석

1 종속변수(D):
우울_2g [GDS15_2g]

2 독립변수(I):
IADL_2g [IADL_2g]
만성질환_2g [ch_dz_2g]
청력 일상생활의 불편함 [C2 2]

통계량(S)...
도표(T)...
저장(S)...
옵션(O)...
부트스트랩(B)...

6 확인

선형 회귀분석: 통계량

회귀계수
☒ 추정값(E)
☐ 신뢰구간(C)
수준(%): 95
☐ 공분산 행렬(V)

☒ 모형 적합(M)
☐ R 제곱 변화량(S)
☐ 기술통계(D)
☐ 부분상관 및 편상관계수(P)
☒ 공선성 진단(L)

잔차
☐ Durbin-Watson(U)
☐ 케이스별 진단(C)
◎ 밖에 나타나는 이상값(O): 3 표준편차
◎ 전체 케이스(A)

5 계속

Multivariable analysis: 다변량(수) 분석

다중 공선성(multicollinearity)

계수^a

모형	비표준화 계수		표준화 계수	t	유의확률	공선성 통계량	
	B	표준오차	베타			공차	VIF
1 (상수)	-.013	.022		-.618	.537		
IADL_2g	.061	.026	.099	2.346	.019	.875	1.142
만성질환_2g	.053	.024	.088	2.180	.030	.947	1.056
청력_일상생활의 불편함	.065	.024	.111	2.755	.006	.957	1.045
인지_2g	.062	.024	.105	2.567	.011	.928	1.077

a. 종속변수: 우울_2g

- VIF 값이 1.05~1.14로 독립변수들 간의 상관성이 높지 않음. → 모든 변수를 최종 모형에 사용할 수 있음.

Multivariable analysis: 다변량(수) 분석

분석-회귀분석-이분로지스틱: 단변량 유의변수 투입

로지스틱 회귀 모형

로지스틱 회귀 모형

종속변수(D): 우울_2g [GDS15_2g]

공변량(C): IADL_2g, ch_dz_2g, C2_2(Cat), cognition_2g(Cat)

방법(M): 앞으로: LR

선택변수(C):

확인, 불여넣기(P), 재설정(R), 취소, 도움말

변수 선택 방법: 앞으로(전진, LR)방법으로!!

×

범주형(C)...
저장(S)...
옵션(O)...
붓스트랩(T)...

로지스틱 회귀분석: 범주형 변수 정의

×

공변량(C):

범주형 공변량(T):

cognition_2g(표시자(처음))
IADL_2g(표시자(처음))
ch_dz_2g(표시자(처음))
C2_2(표시자(처음))

대비 바꾸기

대비(N): 표시자
참조범주(R): ☐ 마지막(L) ☒ 처음(F)

계속, 취소, 도움말

****변수 선택 방법**

- **입력(Enter):** 모든 독립변수를 단일단계의 모형에 **한꺼번에 투입**
- **앞으로(전진선택법, forward selection):** 변수가 모형에 **순차적으로 입력되는** 단계별 변수 선택. 통계량 기준으로 종속변수와 가장 유의한 독립변수부터 하나씩 모형에 추가하여 가장 유의한 독립변수들을 선택
- **뒤로(후진 제거법, backward elimination):** 방정식에 모든 변수를 입력한 다음 **순차적으로 제거**하는 변수 선택법. 통계량을 기준으로 유의하지 않은 독립변수를 제거하면서 유의한 독립변수를 선택함.

Multivariable analysis: 다변량(수) 분석

분석-회귀분석-이분로지스틱: 단변량 유의변수 투입

로지스틱 회귀 모형

로지스틱 회귀 모형

종속변수(D): 우울_2g [GDS15_2g]

블록1대상1

공변량(C): IADL_2g(Cat), ch_dz_2g(Cat), C2_2(Cat), cognition_2g(Cat)

방법(M): 뒤로: LR

선택변수(C):

확인, 불여넣기(P), 재설정(R), 취소, 도움말

범주형(C)...
저장(S)...
옵션(O)...
붓스트랩(T)...

로지스틱 회귀분석: 옵션

로지스틱 회귀분석: 옵션

통계량 및 도표

☒ 분류도표(C)
☒ Hosmer-Lemeshow 적합도(H)
☐ 케이스별 잔차목록(W):
○ 밖에 나타나는 이상값(O) 2 표준편차
○ 전체 케이스(A)
표시
○ 각 단계마다(E) ☒ 마지막 단계에서(L)

☐ 추정값들의 상관계수(R)
☐ 반복계산정보(I)
☒ exp(B)에 대한 신뢰구간(X): 95 %의 케이스 추출

단계선택에 대한 확률
진입(E): 0.05 제거(A): 0.10

☐ 복잡한 분석 또는 큰 데이터 집합을 위해 메모리 보존
☒ 모형에 상수 포함(I)

분류 분리점(U): 0.5
최대반복계산수(M): 20

계속, 취소, 도움말

다변량 이분형 로지스틱 분석 결과: 독립성, 만성질환, 청력불편, 인지 장애 투입

케이스 처리 요약

가중되지 않은 케이스 ^a	N	퍼센트
선택 케이스 분석에 포함	610	100.0
결측 케이스	0	.0
합계	610	100.0
비선택 케이스	0	.0
합계	610	100.0

a. 가중값을 사용하는 경우에는 전체 케이스 수의 분류표를 참조하십시오.

범주형 변수 코딩

		빈도	파라미터 코딩
			(1)
인지_2g	인지저하 없음	280	.000
	인지저하 있음	330	1.000
만성질환_2g	2개이하	372	.000
	3개이상	238	1.000
청력_일상생활의 불편함	불편하지 않음	299	.000
	불편함	311	1.000
IADL_2g	독립	399	.000
	의존	211	1.000

종속변수 코딩

원래 값	내부 값
정상(0-7)	0
우울(8이상)	1

독립변수와 종속변수 범주가 어떻게 코딩 되어있는지 보여줌

블록 0: 시작블록

분류표^{a,b}

감시됨			예측		
			우울_2g		분류정확 %
			정상(0-7)	우울(8이상)	
0 단계	우울_2g	정상(0-7)	552	0	100.0
		우울(8이상)	58	0	.0
전체 퍼센트					90.5

a. 모형에 상수항이 있습니다.

b. 절단값은 .500입니다.

방정식에 포함된 변수

	B	S.E.	Wals	자유도	유의확률	Exp(B)
0 단계 상수항	-2.253	.138	266.440	1	.000	.105

방정식에 포함되지 않은 변수

	점수	자유도	유의확률
0 단계 변수			
IADL_2g(1)	16.359	1	.000
ch_dz_2g(1)	8.612	1	.003
C2_2(1)	13.750	1	.000
cognition_2g(1)	12.226	1	.000
전체 통계량	35.541	4	.000

0단계: 독립변수 제외하고 상수항만 존재하는 단계

- 옵션에서 '마지막 단계에서' 선택하였으므로 최종 모형에 도달하기 까지 각 단계가 모두 보이지 않음(각 단계를 모두 보고 싶다면 '각단계마다' 선택하기!)
- 블록 0과 마지막 단계만 제시

블록 1: 전진 단계선택(우도비)



우도비를 이용한 전진단계법 사용

모형 계수 전체 테스트

		카이제곱	자유도	유의확률
4 단계	단계	4.575	1	.032
	블록	36.421	4	.000
	모형	36.421	4	.000

- 모형계수 전체 테스트(회귀모형의 유의성)

H_0 : 모형으로 설명할 수 없다(모든 독립변수의 계수가 0과 동일함)

H_1 : 모형으로 설명할 수 있다(모형의 $p < .05$ 로 H_1 채택)

→ 추정된 모형의 $p < .001$ 로 추정된 모형은 유의하다고 할 수 있음!

모형 카이제곱(model chi-square)

상수항만을 가진 모형(설명변수가 없는 Null model)의 -2LL값과 K개의 독립변수를 가진 모형(해당 설명변수들을 가진 모형)의 -2LL값의 차이

값이 클수록 모형이 유의미할 확률이 높아짐 → 값이 클수록 적어도 1개 이상의 독립변수가 종속 변수에 영향을 미칠 가능성이 높아짐

<참고>

- 로지스틱회귀분석: -2LL값의 차이를 계산하여 χ^2 계산
- 선형회귀분석: R^2 의 차이를 계산하여 F 검정

모형 요약

단계	-2 Log 우도	Cox와 Snell의 R-제곱	Nagelkerke R-제곱
4	346.830 ^a	.058	.124

a. 모수 추정값이 .001보다 작게 변경되어
계산반복수 6에서 추정을 종료하였습니다.

• 회귀모형 추정방법: 최대우도법 (-2Log Likelihood: 예측치와 관찰치의 비교)

<Log Likelihood(우도)>

- 주어진 추정계수로 관측값이 발생할 확률
- 0~1사이의 값으로 1보다 작음
- 통상적으로 우도에 로그를 취한 뒤 차이 값이 카이제곱분포를 따르게 하기 위해 -2배(-2LL)하여 추정 모델이 자료에 얼마나 적합하지 측정하는 지표로 사용

로지스틱 회귀분석에서 좋은 모형은 관찰 결과가 나올 우도가 높은 모형(**-2LL가 작은 모형**)

모형이 자료와 완벽하게 일치한다면, -2LL=0!!

• 회귀 모형의 유의성: Cox&Snell 과 Nagelkerke의 결정계수

본 분석의 회귀식에서 설명되는 변이의 추정치(Nagelkerke는 Cox&Snell의 수정 설명력으로 1이 될 수 없음. 주로 **Nagelkerke의 결정계수 사용!!**)

→ Nagelkerke의 결정계수(R^2)가 0.124 이므로 회귀식은 현상을 12.4% 설명할 수 있다.

다변량 이분형 로지스틱 분석 결과: 독립성, 만성질환, 청력불편, 인지 장애 투입

= Hosmer 와 Lemeshow 검정 =

단계	카이제곱	자유도	유의확률
4	1.270	6	.973

Hesmer와 Lemeshow 검정에 대한 분할표

		우울_2g = 정상(0-7)		우울_2g = 우울(8이상)		합계
		감시됨	예상됨	감시됨	예상됨	
4 단계	1	154	153.219	4	4.781	158
	2	60	60.675	4	3.325	64
	3	37	36.377	2	2.623	39
	4	105	106.623	10	8.377	115
	5	33	33.729	5	4.271	38
	6	56	54.483	6	7.517	62
	7	51	50.781	9	9.219	60
	8	56	56.112	18	17.888	74

분류표^a

감시됨			예측		
			우울_2g		분류정확 %
			정상(0-7)	우울(8이상)	
4 단계	우울_2g	정상(0-7)	552	0	100.0
		우울(8이상)	58	0	.0
전 체 퍼센트					90.5

a. 절단값은 .500입니다.

적합도 검정에서는 자료가 모형에 적합하다고 평가되었으나 분류표에서는 정상과 우울을 명확히 구분하지 못하고 있음. 전체적인 분류 정확도는 90.5% 임.

- **Goodness-of-fit test**(적합성 통계량)

→ **Hosmer Lemeshow 검정으로 확인**

H_0 : 모형이 적합하다

H_1 : 모형이 적합하지 않다.

최종 4단계 모형의 $p=.973(>0.05)$ 이므로 H_0 채택 → **자료가 모형에 적합하다고 평가됨.**

<알아두기!!>

로지스틱 회귀분석으로 추정된 모형의 평가

- 모형계수 전체 테스트(Model chi-square)
- 분류표 확인
- Cox&Snell 과 **Negelkerke 결정계수**로 설명력확인
- Hosmer-Lemeshow 검정 확인

→ 하나의 절대적 방법은 없음

→ 여러 결과들로 종합적으로 판단필요

다변량 이분형 로지스틱 분석 결과: 독립성, 만성질환, 청력불편, 인지 장애 투입

방정식에 포함된 변수

	B	S.E.	Wals	자유도	유의확률	Exp(B)	EXP(B)에 대한 95% 신뢰구간	
							하한	상한
4 단계 ^a IADL_2g(1)	.649	.301	4.643	1	.031	1.913	1.060	3.452
ch_dz_2g(1)	.625	.293	4.543	1	.033	1.868	1.052	3.319
C2_2(1)	.878	.320	7.525	1	.006	2.405	1.285	4.503
cognition_2g(1)	.856	.332	6.657	1	.010	2.354	1.229	4.512
상수항	-3.936	.387	103.647	1	.000	.020		

a. 변수가 4: 단계에 진입했습니다 ch_dz_2g. ch_dz_2g.

- **B:** 회귀계수
- **Wals:** 각 회귀계수의 유의성 검정=(회귀계수의 추정치/표준오차)²
회귀계수 추정치의 절대값이 크면 클수록 Wald 검정값이 기하급수적으로 커짐 → 회귀계수가 커도 가설검정을 위해 Wald 값을 너무 신뢰해서는 안됨.
- **Exp(B): 승산비**
 - 1 초과: 정(+)의 영향
 - 1 미만: 부(-)의 영향
- 우울의 예측요인은 독립성(IADL_2g), 만성질환(ch_dz_2g), 청력불편(C2_2), 인지장애(cognition_2)임.
- 유의확률은 모두 0.05 미만으로 유의함
- 일상생활에 의존적인 노인이(1) 독립적(0)인 노인에 비해 우울이 1.91배 높음
- 만성질환이 3개 이상(1)인 노인이 2개 이하(0)인 노인에 비해 우울이 1.87배 높음
- 청력 불편감이 있는(1)노인이 없는 노인(0)에 비해 우울이 2.41배 높음
- 인지저하가 있는 노인(1)이 없는 노인(0)에 비해 우울이 2.35배 높음

Multivariable analysis table

다변량 분석에서 OR
=Adjusted OR

Variable	OR	95%CI	<i>p</i>
IADL_2g(독립성)			
ch_dz_total(만성질환)			
C1_3(청력불편감)			
Cognition_2g(인지장애)			
Constant(상수항)			
Hosmer-Lemeshow tset: $\chi^2=$, $df=$, $p=$ Model summary: Nagelkerke R2=			

Tip

- 범주형 독립변수의 OR값이 이상한 경우(ex. OR=30)
 - 교차비를 꼭 확인
 - 일부 범주의 빈도가 작거나 0인 경우에 OR 값이 비정상적으로 높게 나올 수 있음
 - 다른 범주와 결합하여 분석진행 필요

Thank You.

Q&A



한양대학교간호과학연구소
Research Institute of Nursing Science Hanyang University

