

11th Week

Bivariate Regression Analysis

이변량 회귀분석

2011

Jongseok Lee

Business Administration

Hallym University

주식 가격 = $f(?)$



Source: <http://finance.naver.com>

변수들 관계의 모형화: 확정적 모형과 확률적 모형

확정적 모형 Deterministic Model

오차항이 없다!

- 직사각형의 면적 = 가로 $\times$ 세로
- 원의 면적 = $3.14 \times$ 반지름의 제곱
- 몸무게 = 체지방 + 제지방

확률적 모형 Probabilistic Model

오차항이 있다!

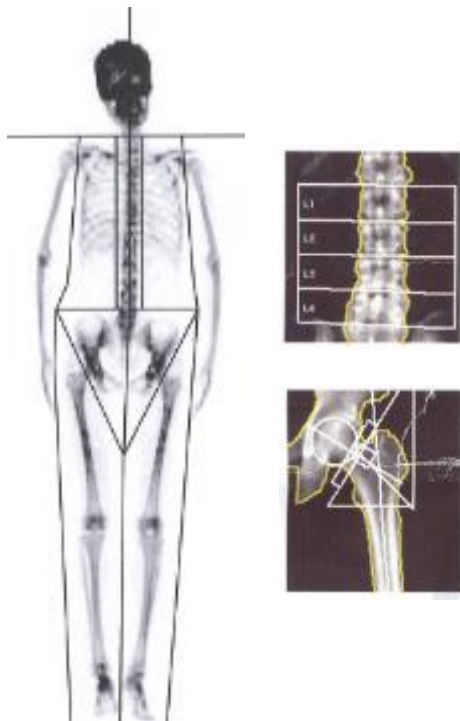
□ 몸무게 = $f(\text{키}) + \epsilon$

ϵ : 입실론(epsilon) = 오차항(error term) 혹은 무작위항(random term)

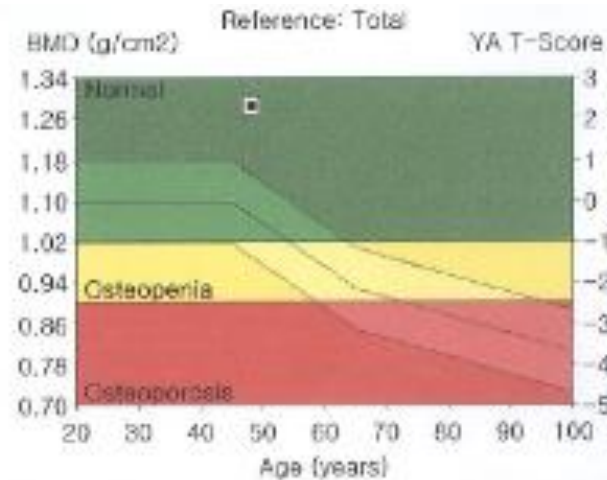
☞ 실제 몸무게와 키를 근거로 추정된 몸무게의 차이

□ 몸무게 = $f(\text{키}, \text{허리둘레}, \dots) + \epsilon$

$$\text{골밀도} = f(?) + \epsilon$$



1.284 g/cm²



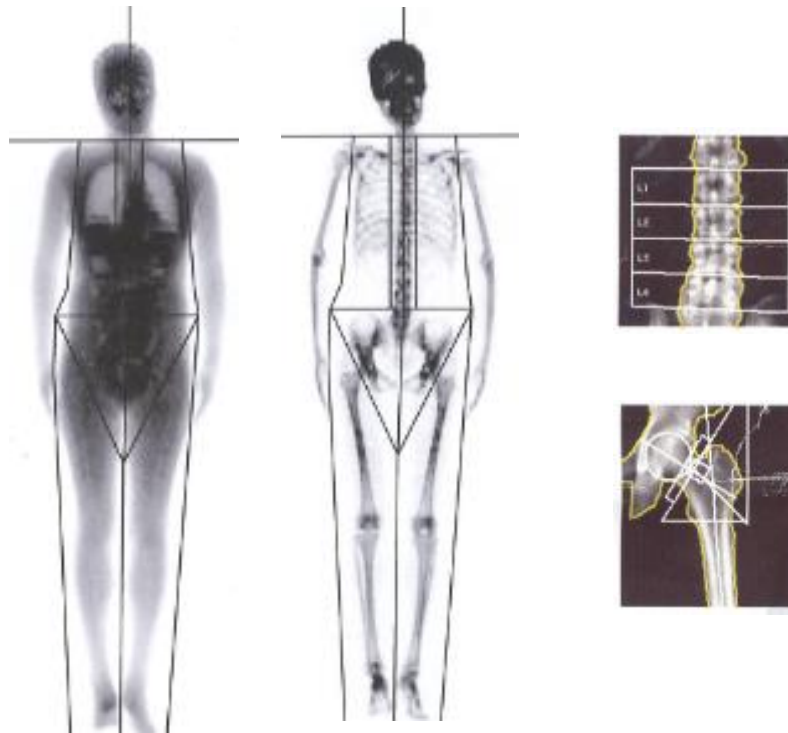
Region	BMD (g/cm ²)	1		2		3	
		Young-Adult (%)	T-Score	Age-Matched (%)	Z-Score		
Head	2.615	-	-	-	-		
Arms	1.018	-	-	-	-		
Legs	1.181	-	-	-	-		
Trunk	1.073	-	-	-	-		
Ribs	0.773	-	-	-	-		
Pelvis	1.290	-	-	-	-		
Spine	1.279	-	-	-	-		
Total	1.284	117	2.3	120	2.7		

Act 1

골밀도 알아 맞추기

1.284 g/cm²

$$= \text{BMC} / \text{Area} = 2,557 \text{ (g)} / 1,991(\text{cm}^2)$$



전체의 골밀도 **평균**으로 추정

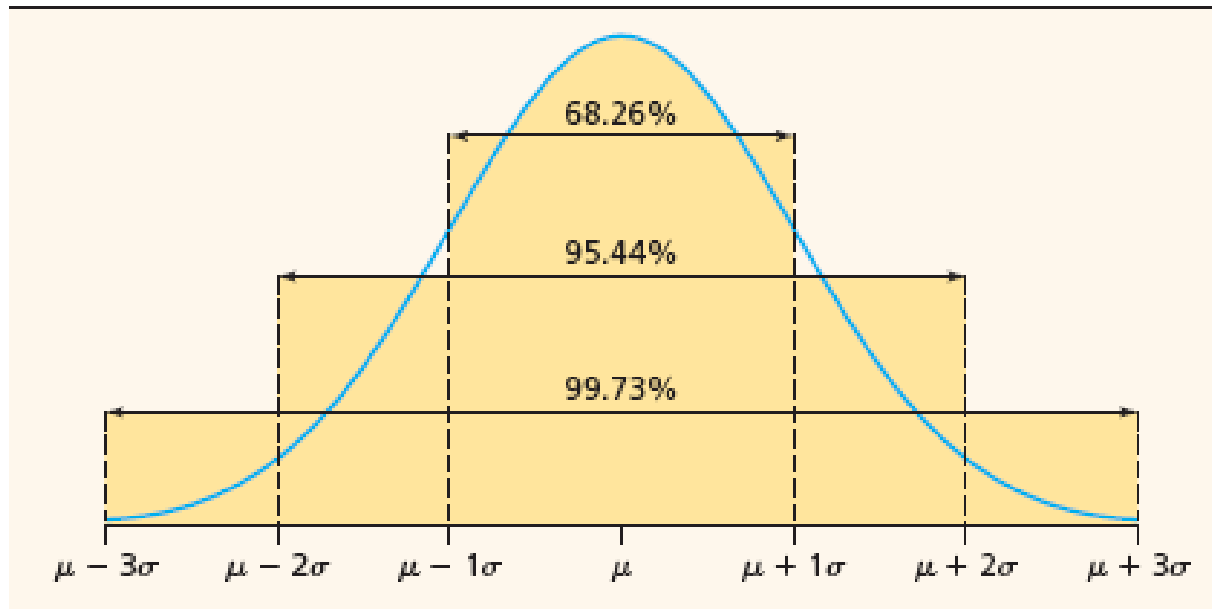


상대에 대한 정보가 하나도 없는데
골밀도를 어떻게 맞추지?

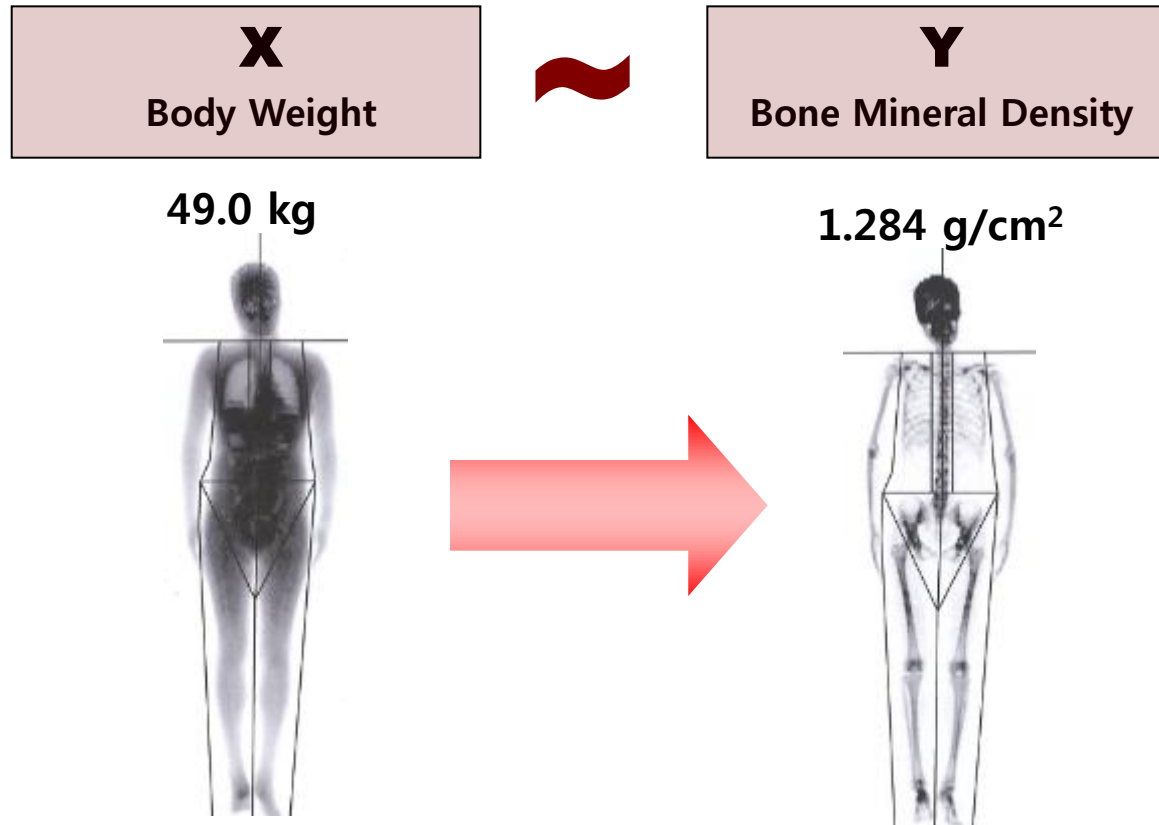


정규 분포

Normal Distribution



Weight and BMD



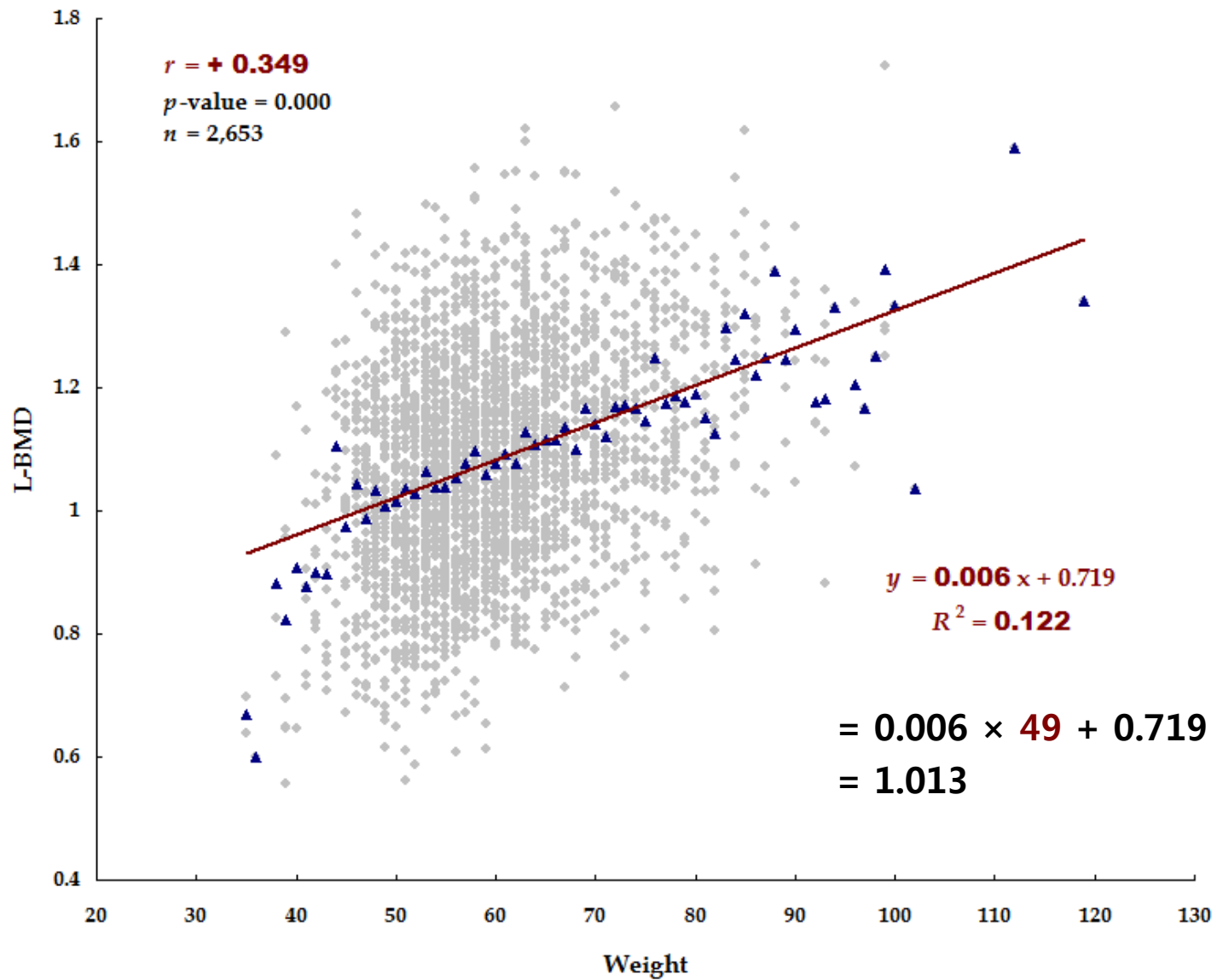
BMD is associated with body weight !

체중이 49 kg인 사람들의 골밀도 평균으로 추정

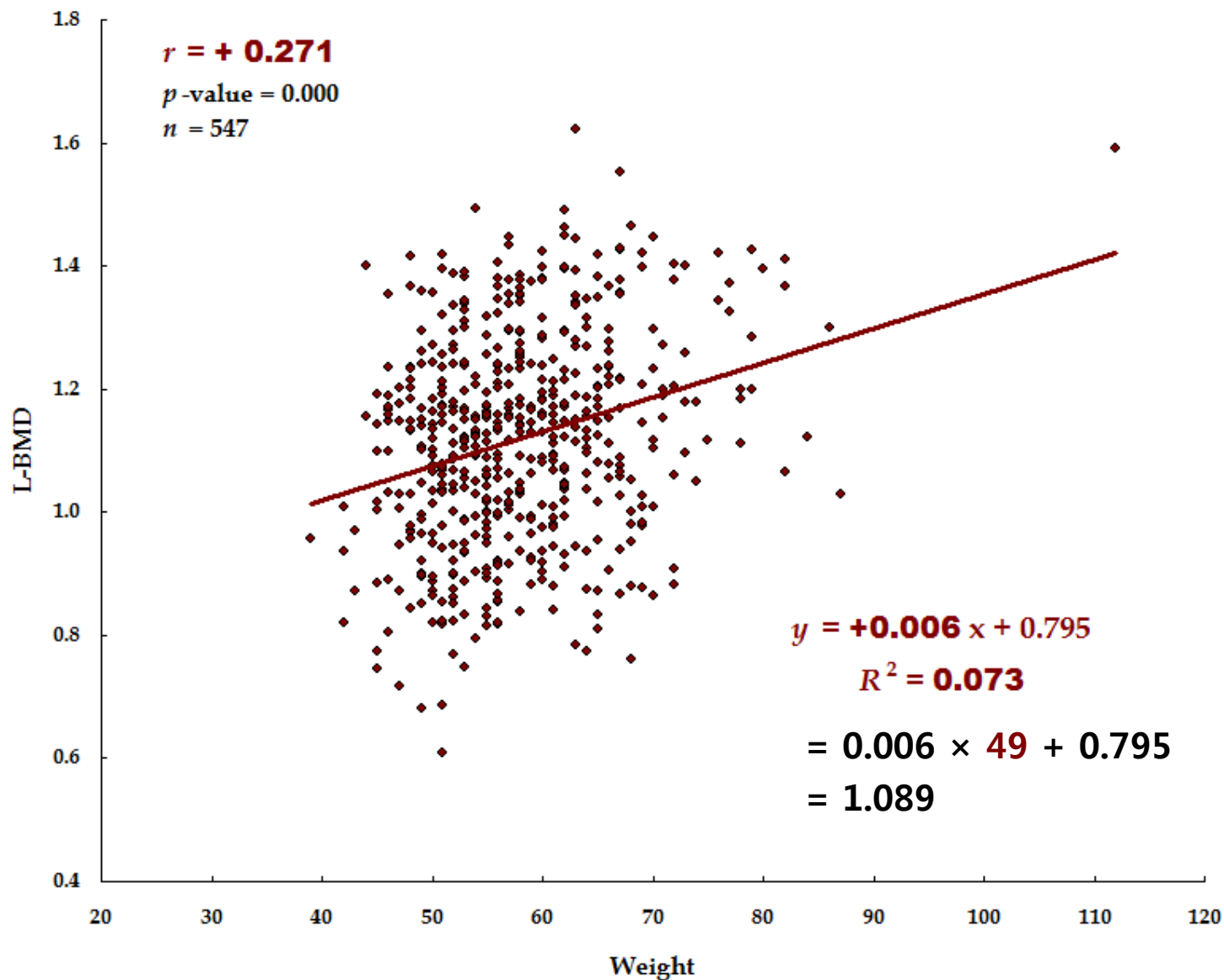
체중이 49 kg이에요.



체중에 관한 정보만 있는 경우



체중뿐 아니라 여자라는 것을 안 경우



X
Body Weight



Y
Bone Mineral Density

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

선형모형 (linear model)

$$y = b_0 + b_1 x$$

49.0 kg



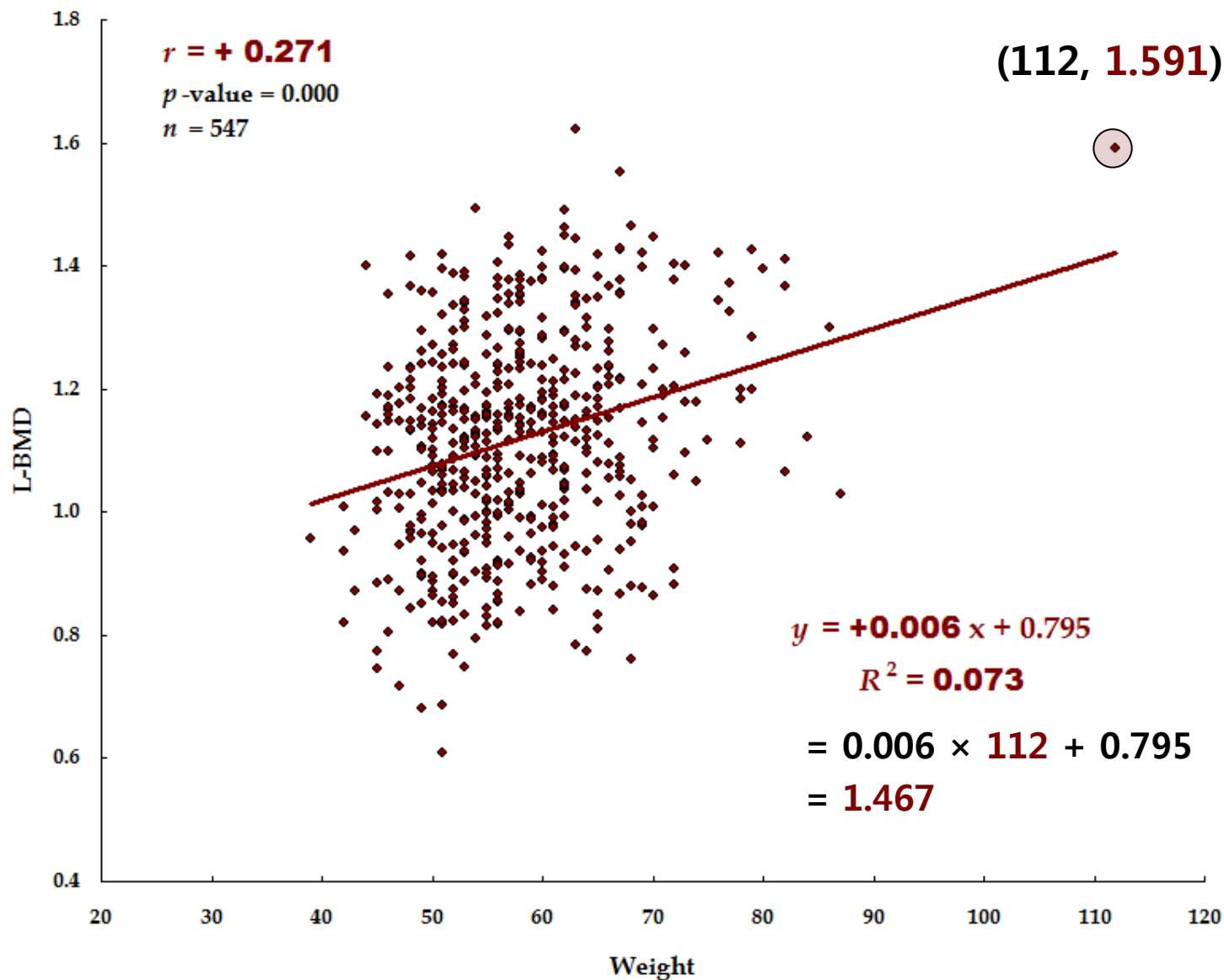
1.284 g/cm²

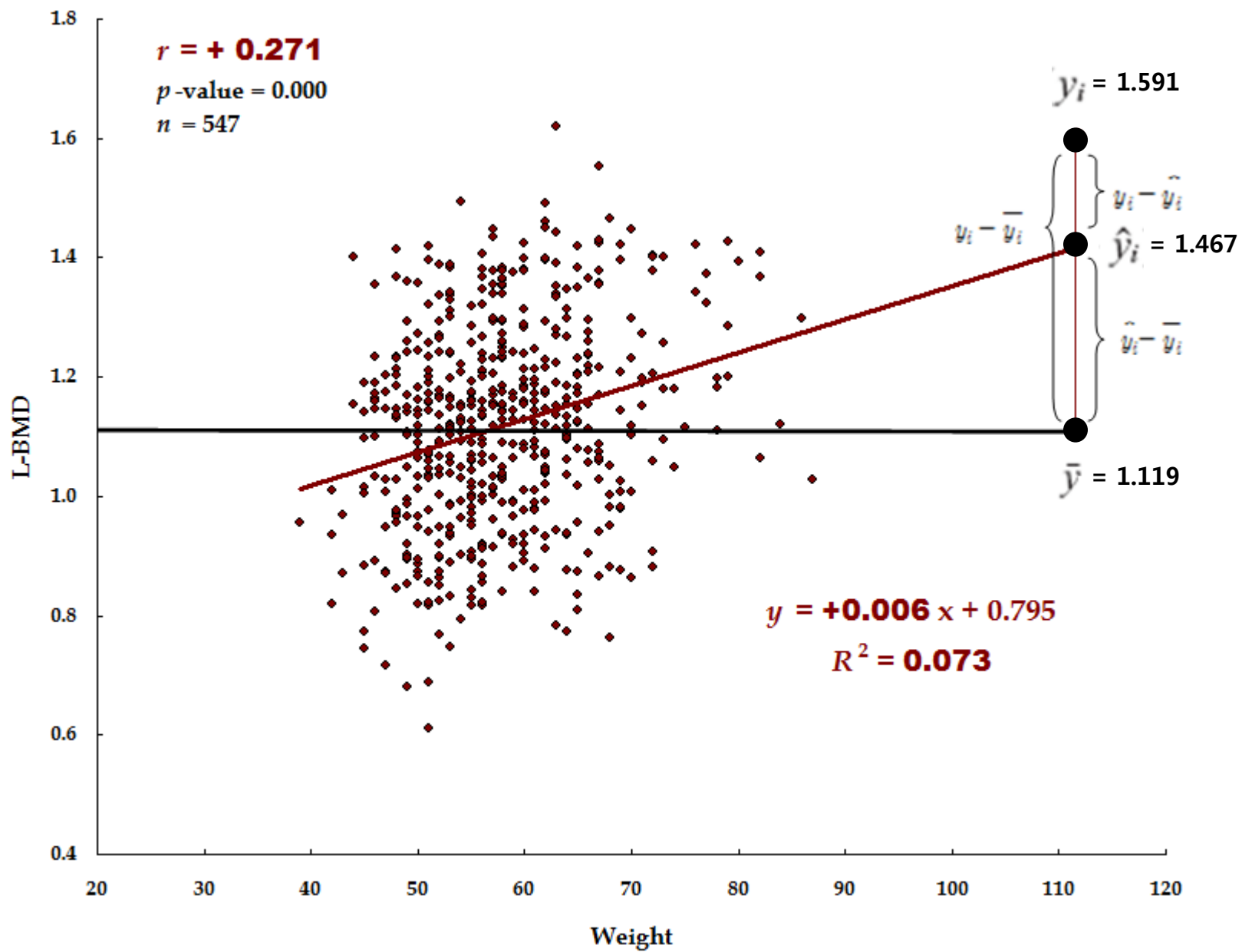


$$\begin{aligned} y &= 0.795 + 0.006 x \\ &= 0.795 + 0.006 \times 49 \\ &= 1.089 \end{aligned}$$

$$\begin{aligned} \varepsilon &= 1.284 - 1.089 \\ &= 0.195 \end{aligned}$$

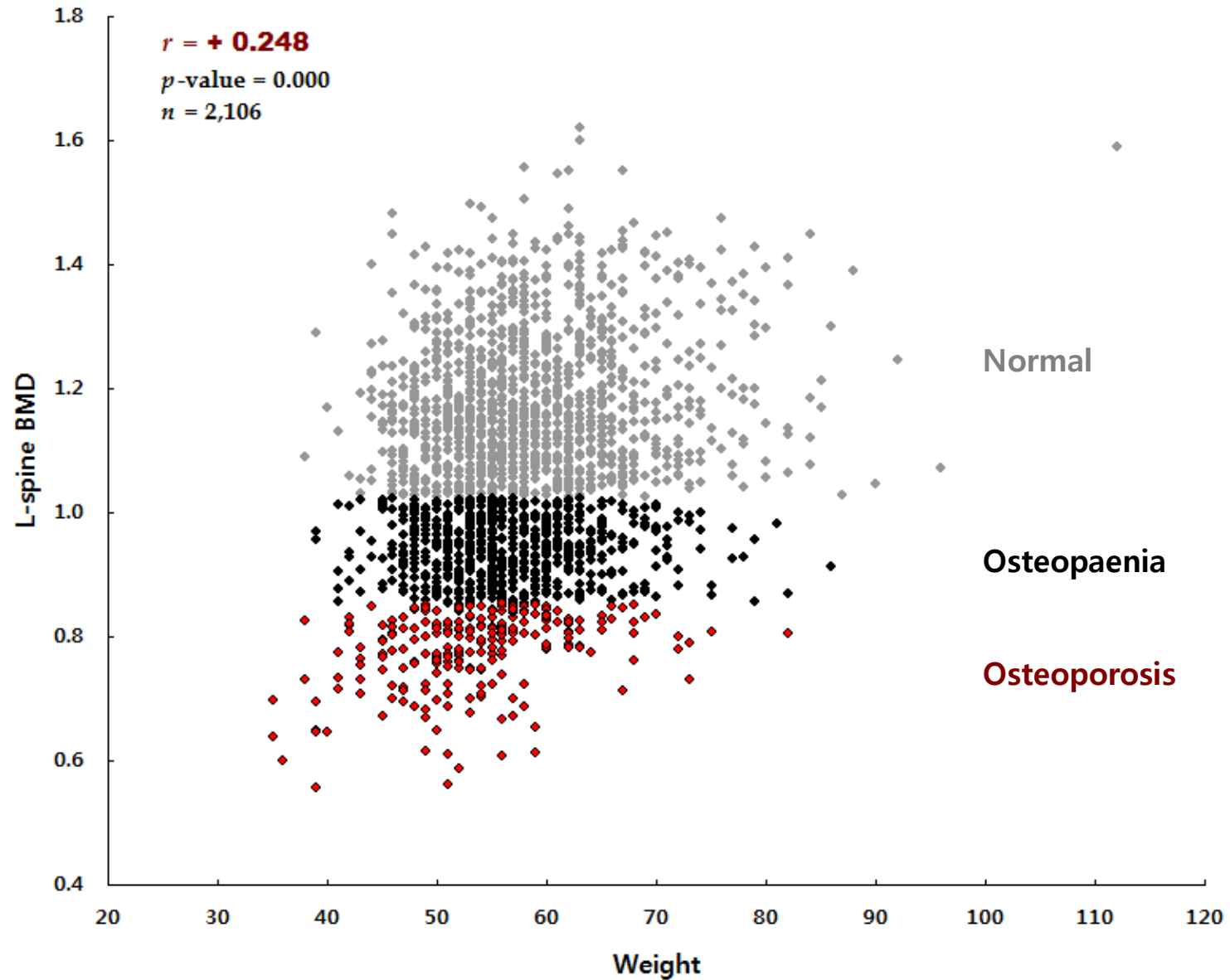
체중이 112 kg인 여자의 골밀도 추정치는?

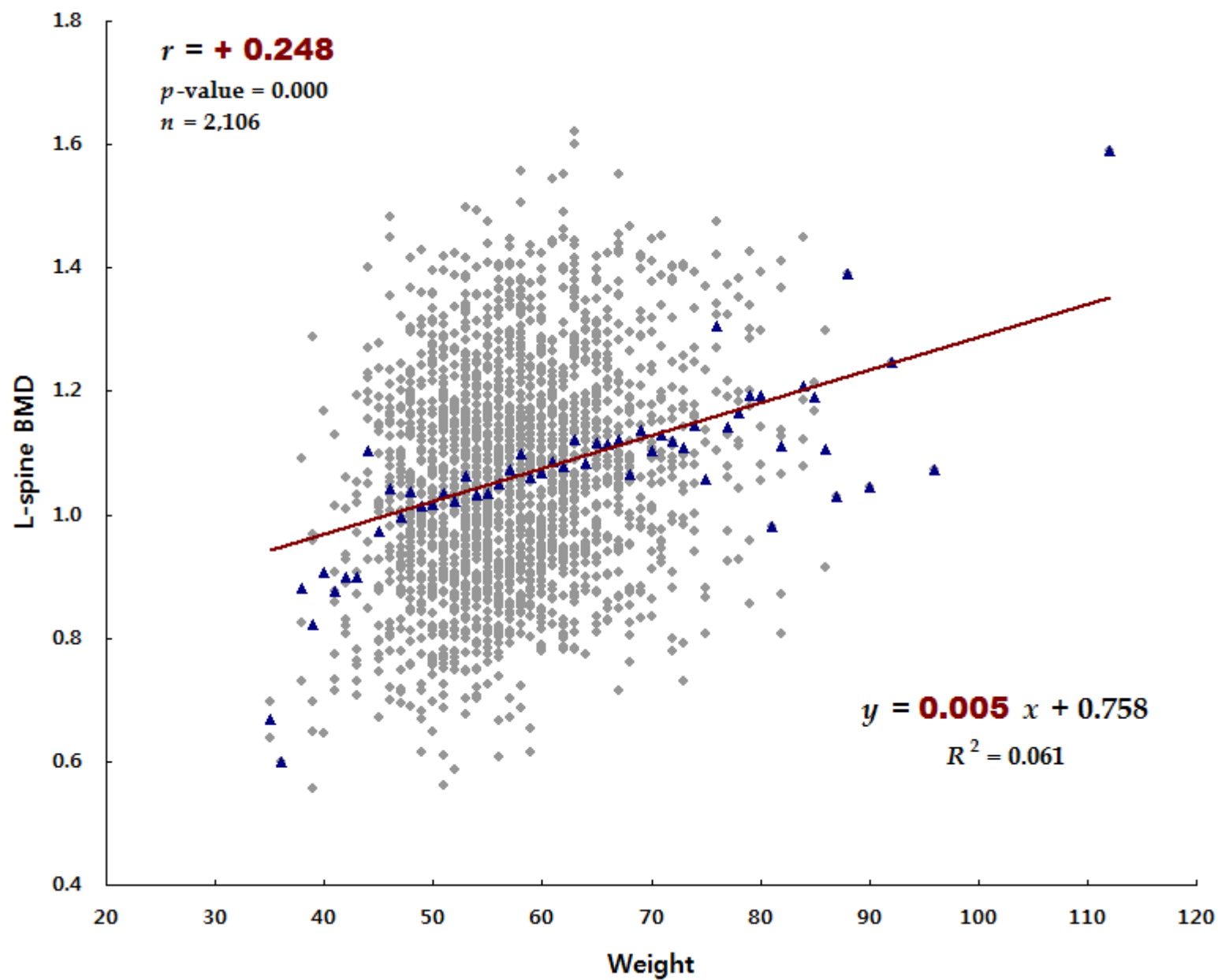




Act 2

Relationship between Weight and BMD





SPSS를 이용한 회귀 분석

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

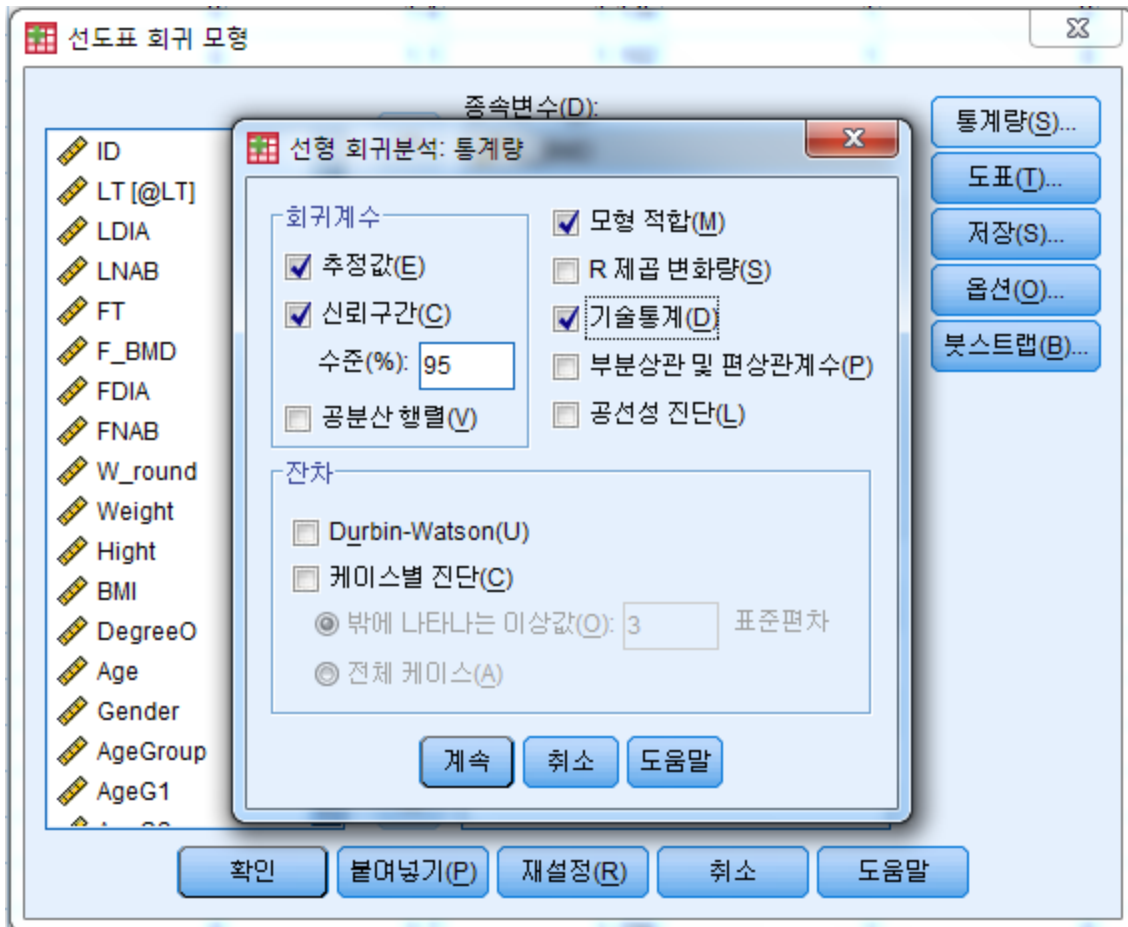
1: ID 1

	ID	@LT
1	1	-
2	2	-
3	3	1.
4	4	-
5	5	-1.
6	6	-2.
7	7	-1.
8	8	-2.
9	9	-1.
10	10	1.
11	11	-
12	12	-
13	13	-2.
14	14	-
15	15	1.
16	16	2.
17	17	-1.
18	18	-
19	19	-
20	20	-
21	21	-
22	22	1

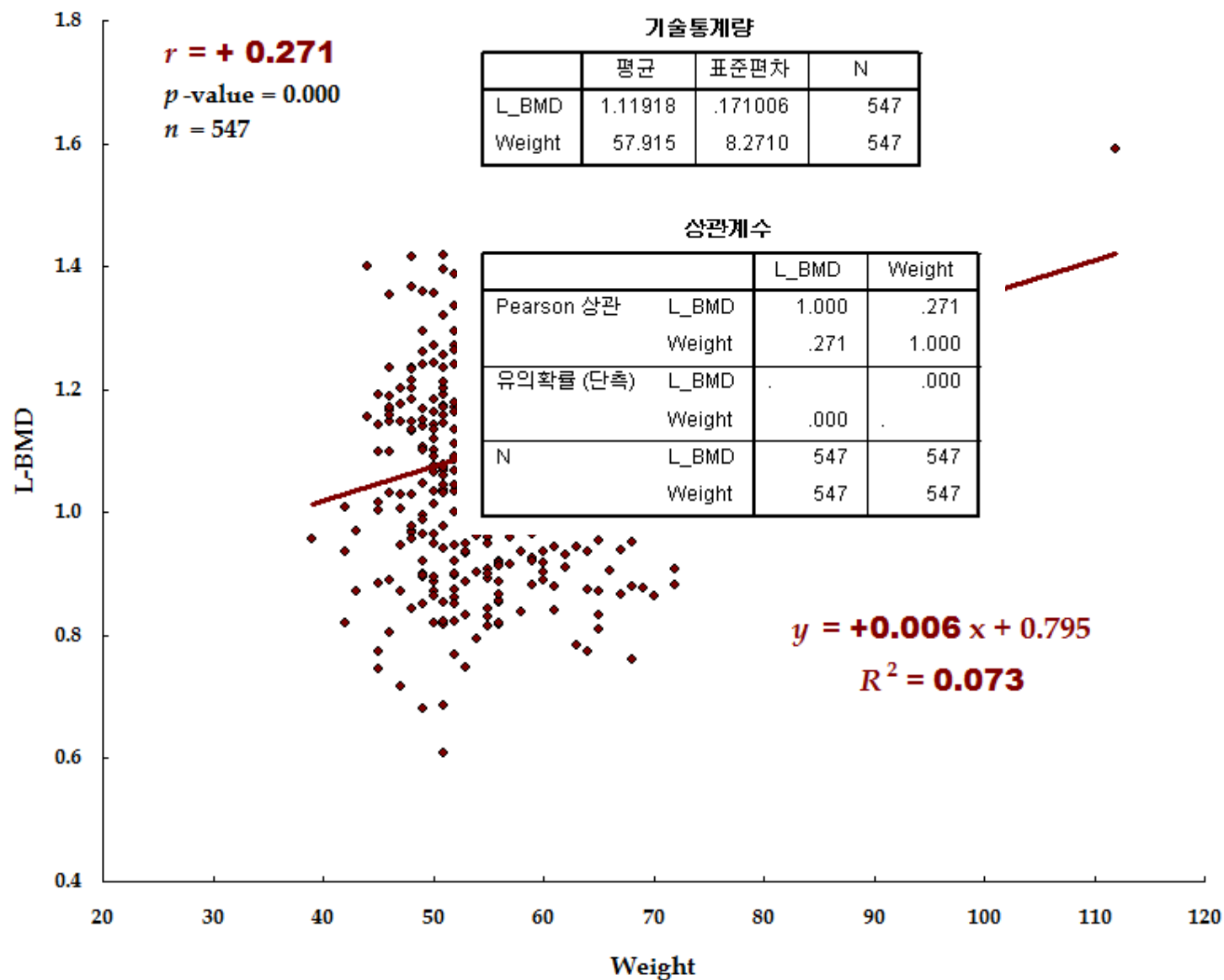
보고서(P)
기술통계량(E)
표
평균 비교(M)
일반선형모형(G)
일반화 선형 모형(Z)
혼합 모형(X)
상관분석(C)
회귀분석(R)
로그선형분석(O)
신경망(W)
분류분석(Y)
차원 감소(D)
척도(A)
비모수 검정(N)
예측(T)
생존확률(S)
다중응답(U)
결측값 분석(V)...
다중 대입(T)
복합 표본(L)
품질 관리(Q)
ROC 곡선(V)...

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)...

DIA	LNAB	FT	F_BMD
1	0	-6	.906
1	0	-3	.942
1	0	1.4	1.136
1	0	1.1	1.102
2	1	-1	.966
		-1.4	.809
		-1.5	.797
		.1	.984
		.5	1.032
		.2	.997
		-.2	.947
		-2.0	.732
		-.5	.918
		1.4	1.147
		2.5	1.275
		-1.4	.807
1	0	1.6	1.165
1	0	-.7	.886
1	0	-.2	.949
1	0	.5	1.039



SPSS 회귀 분석 결과



모형 요약

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차
1	.271 ^a	.073	.072	.164778

a. 예측값: (상수), Weight

분산분석^b

모형	제곱합	자유도	평균 제곱	F	유의확률
1 회귀 모형	1.169	1	1.169	43.054	.000 ^a
잔차	14.798	545	.027		
합계	15.967	546			

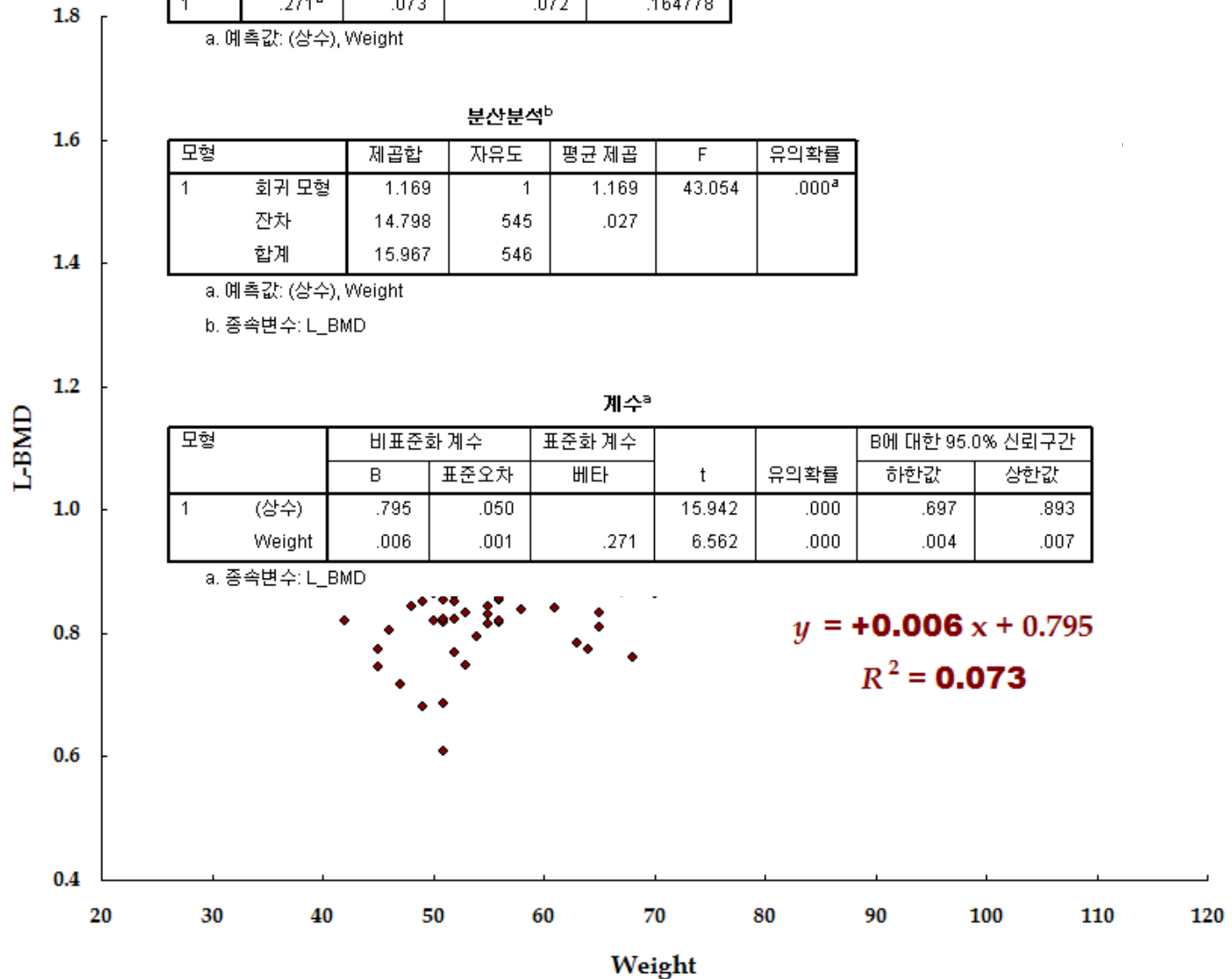
a. 예측값: (상수), Weight

b. 종속변수: L_BMD

계수^a

모형	비표준화 계수		표준화 계수	t	유의확률	B에 대한 95.0% 신뢰구간	
	B	표준오차	베타			하한값	상한값
1 (상수)	.795	.050		15.942	.000	.697	.893
Weight	.006	.001	.271	6.562	.000	.004	.007

a. 종속변수: L_BMD



분석결과의 제시

Associations of L-BMD with Weight by Bivariate Linear Regression Analysis

Dependent Variable	Independent Variable	Coefficient	Standard Error	p-value	R ²
L-BMD	Intercept	0.795	0.050	0.000	0.073**
	Weight	0.006	0.001	0.000	(0.000)

** : p-value < 0.01 * : p-value < 0.05 by ANOVA analysis

모형 요약

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차
1	.271 ^a	.073	.072	.164778

a. 예측값: (상수), Weight

분산분석^b

모형	제곱합	자유도	평균 제곱	F	유의확률
1 회귀 모형	1.169	1	1.169	43.054	.000 ^a
잔차	14.798	545	.027		
합계	15.967	546			

a. 예측값: (상수), Weight

b. 종속변수: L_BMD

계수^a

모형	비표준화 계수		표준화 계수	t	유의확률	B에 대한 95.0% 신뢰구간	
	B	표준오차	베타			하한값	상한값
1 (상수)	.795	.050		15.942	.000	.697	.893
Weight	.006	.001	.271	6.562	.000	.004	.007

a. 종속변수: L_BMD

Association of L-BMD T-Score with Age

Dependent Variable	Independent Variable	Coefficient	Standard Error	p-value	R ²
L-BMD T-Score	Age	-0.055	0.005	0.000	0.143**

** : p-value < 0.01 * : p-value < 0.05 by bivariate linear regression analysis

모형 요약

모형	R	R 제곱	수정된 R 제곱	추정값의 표준오차
1	.378 ^a	.143	.142	1.2228

a. 예측값: (상수), Age

분산분석^b

모형		제곱합	자유도	평균 제곱	F	유의확률
1	회귀 모형	203.958	1	203.958	136.576	.000 ^a
	잔차	1223.063	819	1.493		
	합계	1427.020	820			

a. 예측값: (상수), Age

b. 종속변수: LBMD

계수^a

모형		비표준화 계수		표준화 계수	t	유의확률
		B	표준오차	베타		
1	(상수)	2.470	.245		10.075	.000
	Age	-.055	.005	-.378	-11.687	.000

a. 종속변수: LBMD

체중과 골밀도

Weight and BMD

저체중은 골다공증의 원인으로 알려져 있다. 체중은 골밀도의 좋은 설명변수인가? 골밀도 (T-Score)의 체중(kg)에 대한 이변량 회귀분석의 결과는 다음과 같다.

Association of L-BMD T-Score with Weight

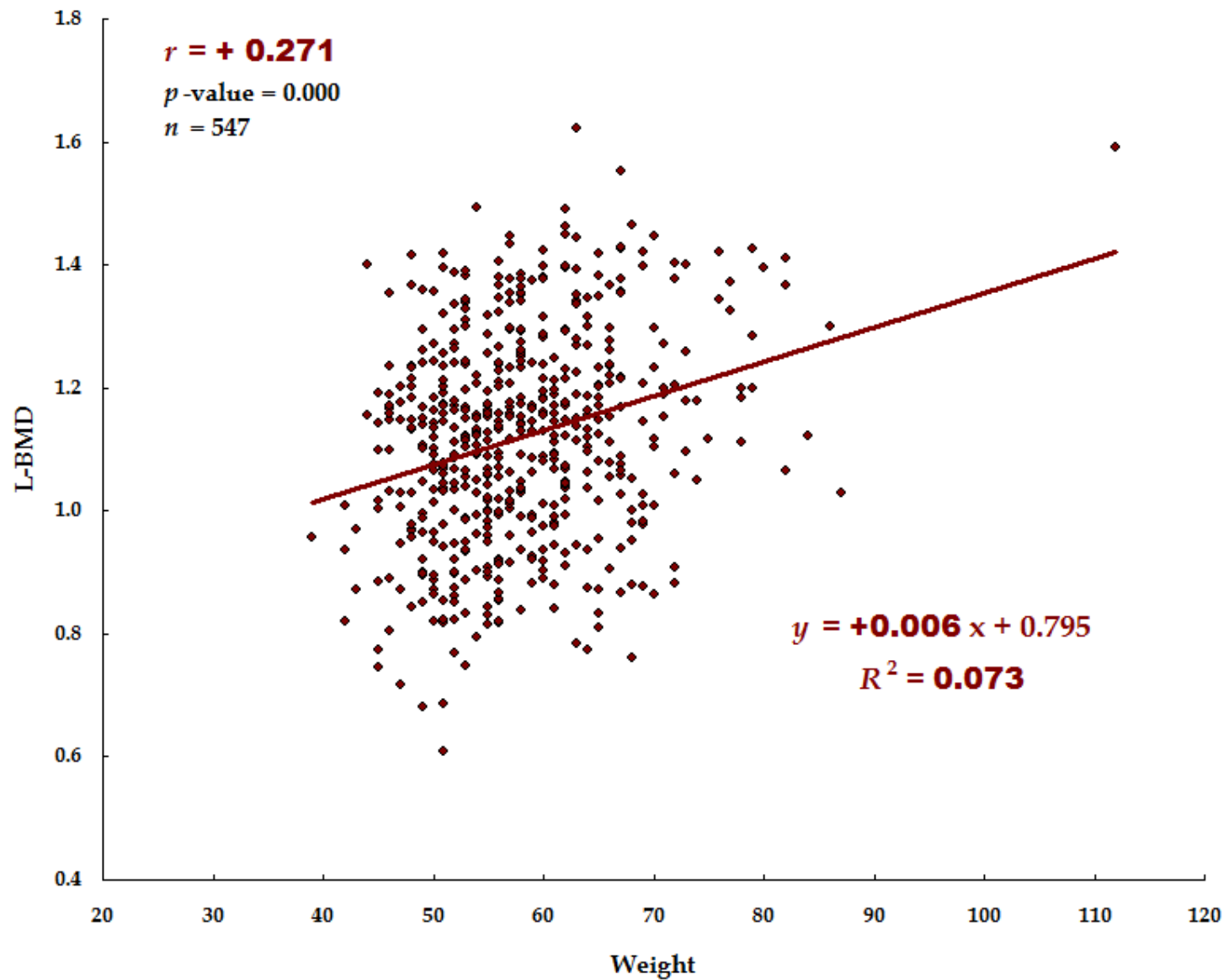
Dependent Variable	Independent Variable	Coefficient	Standard Error	p-value	R ²
L-BMD T-Score	Weight	0.033	0.004	0.000	0.061**

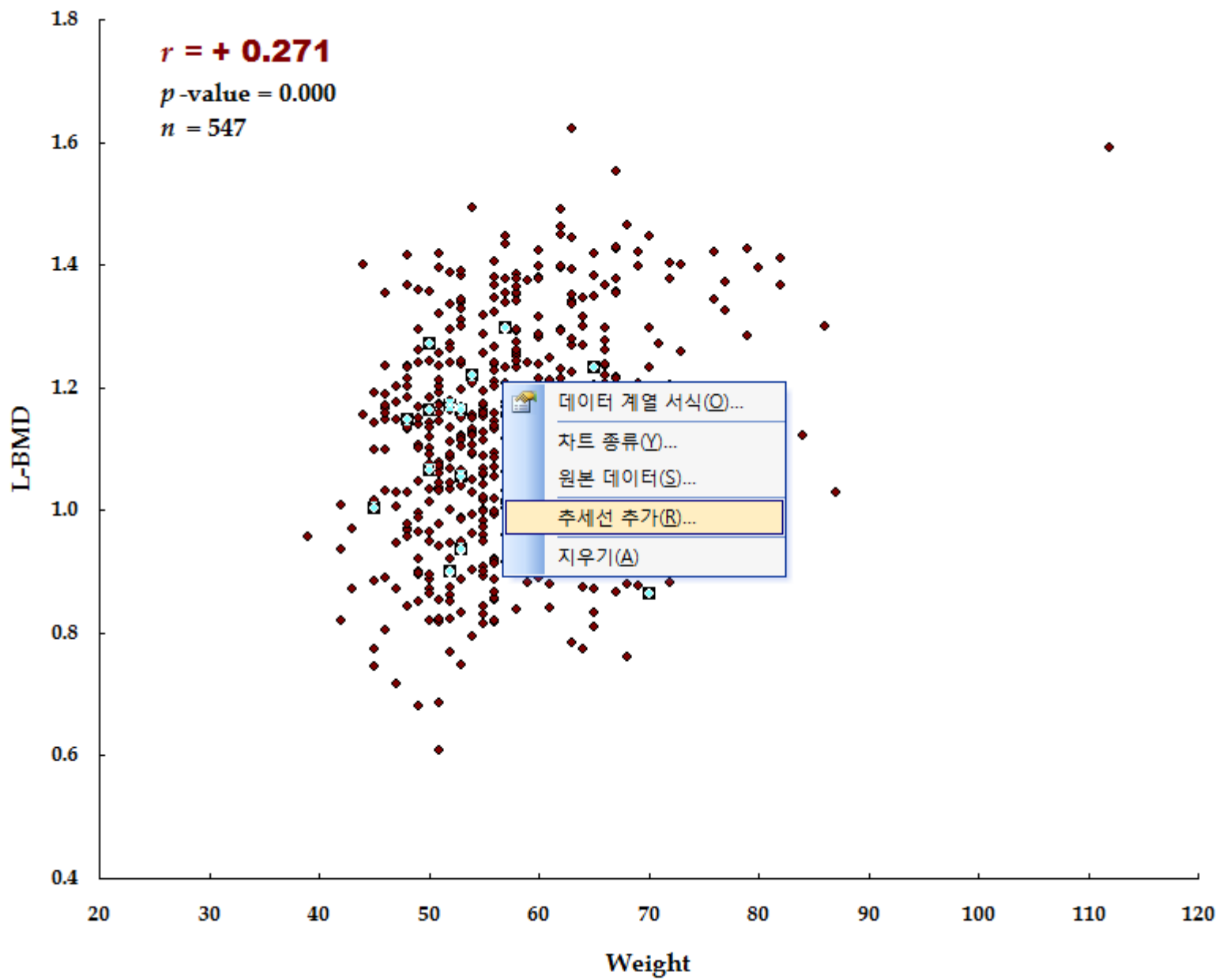
** : p-value < 0.01 * : p-value < 0.05 by bivariate linear regression analysis

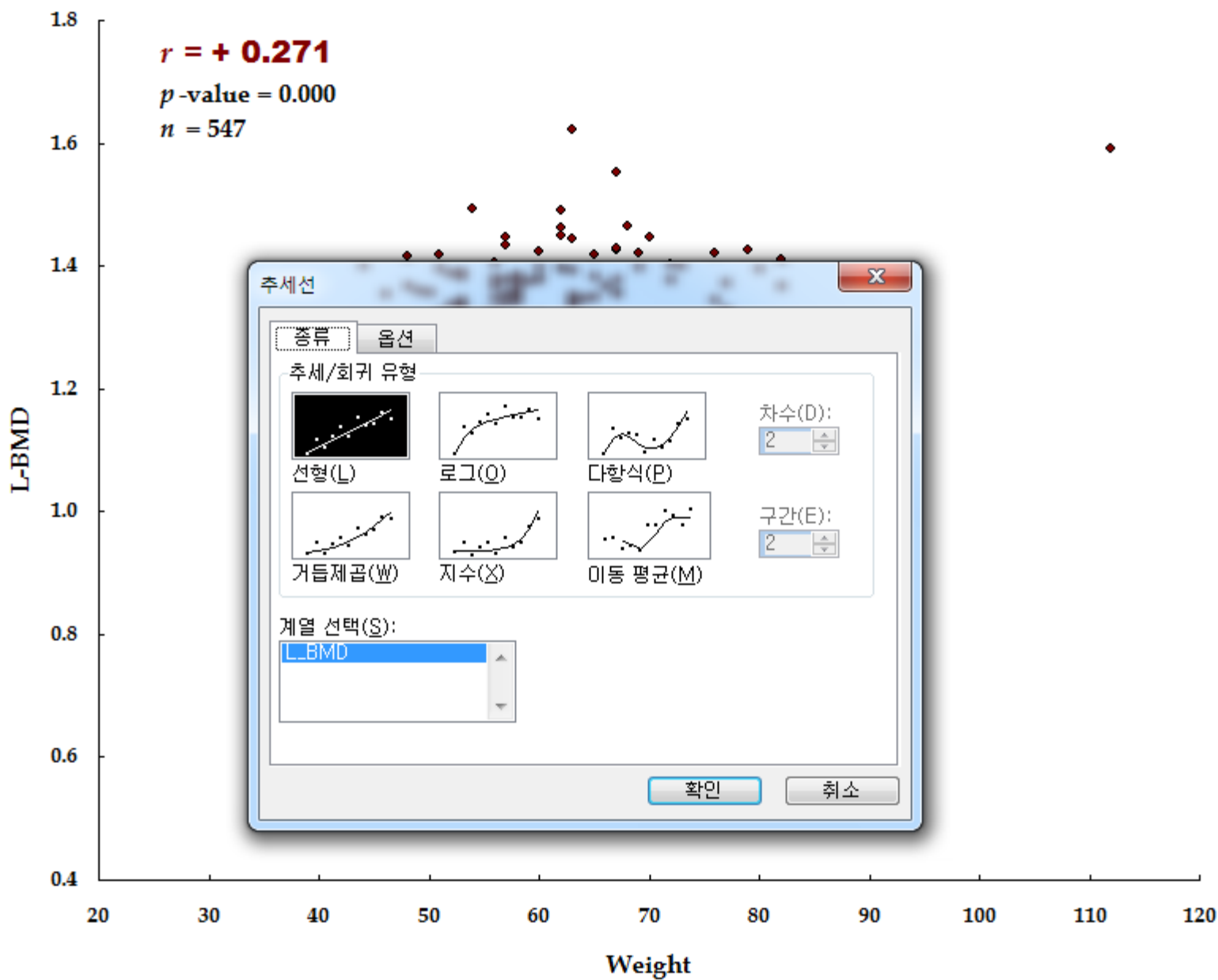
- 적합된 회귀직선의 골밀도(T-Score)의 변동에 대한 설명력은 6.1% 정도 이었다. 회귀모형의 유의성에 대한 분산분석의 결과, 유의수준 0.01에서 적합된 회귀모형이 유의하지 않다는 귀무가설을 기각할 수 있었다.
- 체중과 관련된 회귀계수는 0.033이 되어 체중이 1(kg) 증가하면 골밀도 T값이 0.033 개선되는 것으로 추정되었다. 회귀계수에 대한 유의성 검정 결과, p값이 0.000이 되어 추정된 회귀계수가 유의하지 않다는 귀무가설을 유의수준 0.001에서 기각할 수 있었다.

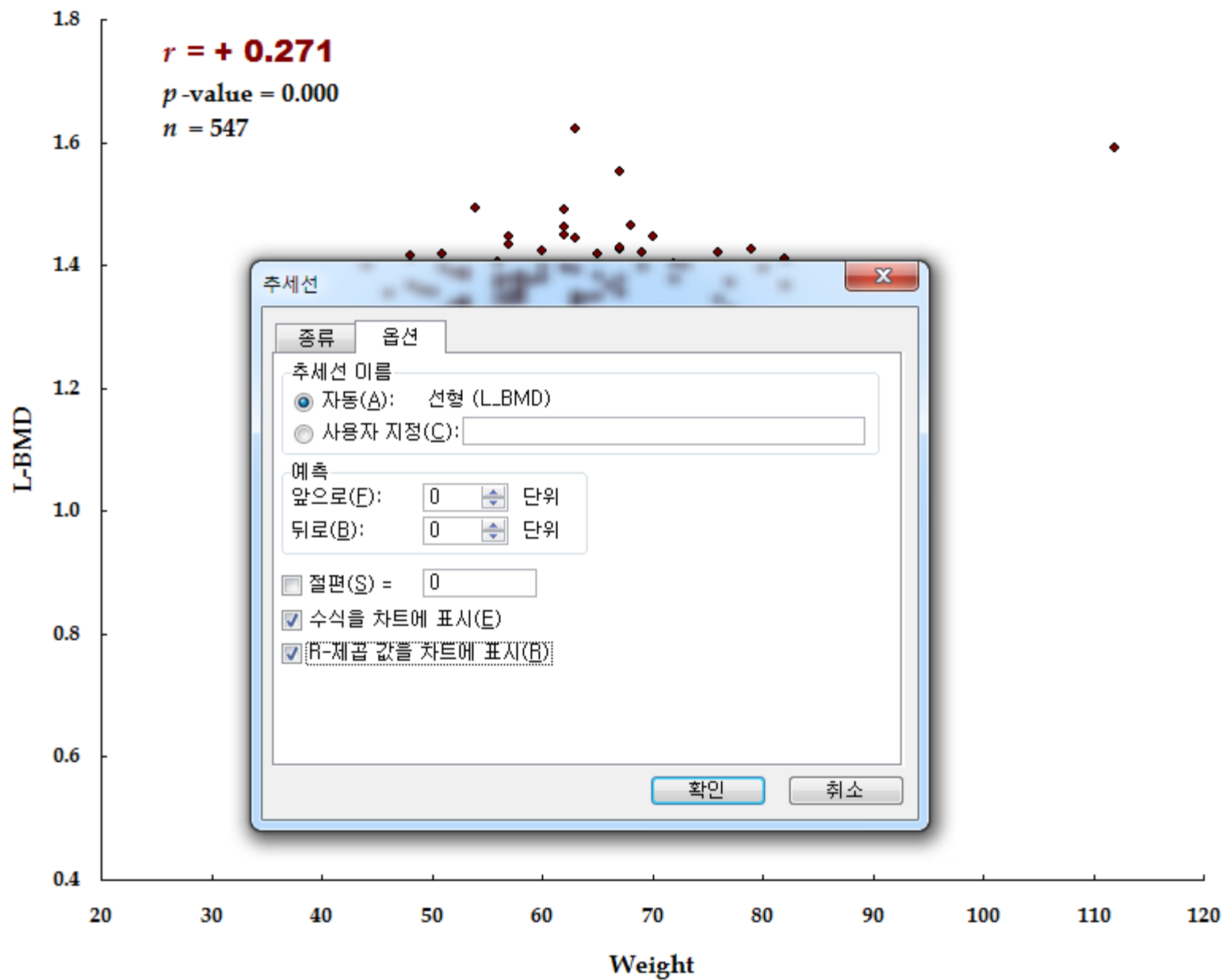
Act 3

Is BMD associated with body weight ?

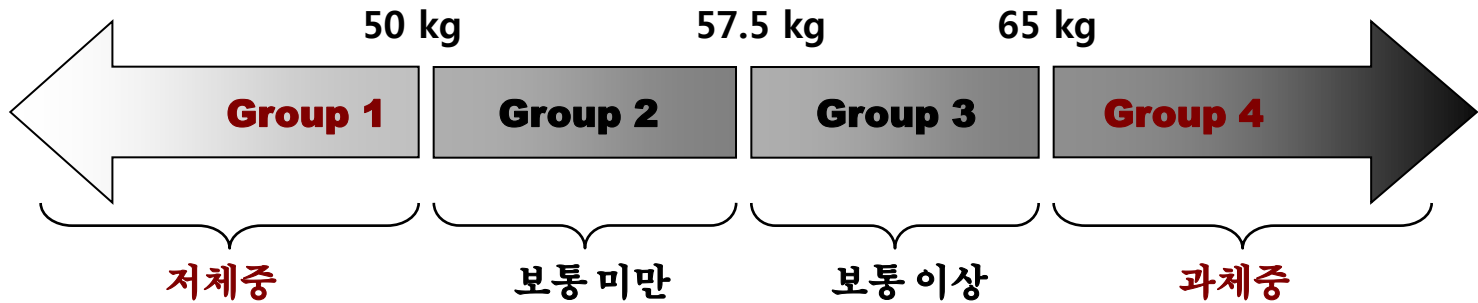








Body Weight



체중 집단들의 적합된 회귀식의 기울기에
통계적으로 유의한 차이가 있는가?

가설 : 체중증가가 골밀도를 향상시킨다면,
과체중 집단의 기울기는 저체중 집단의 기울기 보다
완만할 것이다.

Relationship between Age and BMD

Table. Simple Regression Results

Independent Variable	β (standardized)	S.E.	t-value	p-value	95% CI	R^2 (p-value)
for total (n=2,106)						0.332 (0.000)
Age	-0.010 (-0.577)	0.000	-32.372	0.000	(-0.010, -0.009)	
Intercept	1.584	0.016	96.530	0.000	(1.551, 1.616)	

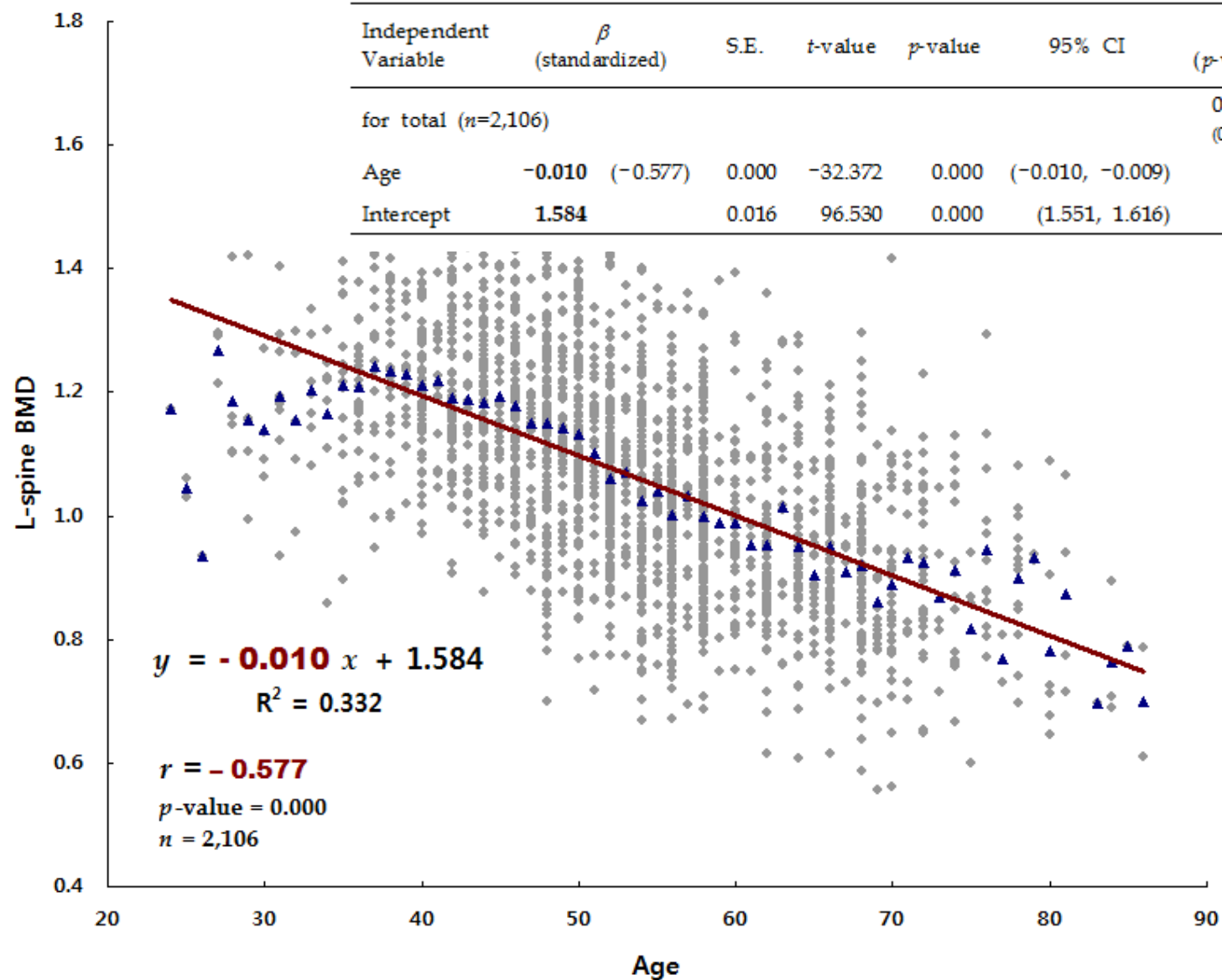


Table. Results of Bivariate Linear Regression Analysis with Age, as an Independent Variable, by Each Weight Group

Independent Variable	β (standardized)	S.E.	t-value	p-value	95% CI	R^2 (p-value)
Total (n=2,106)						0.332 (0.000)
Age	-0.010 (-0.577)	0.000	-32.372	0.000	(-0.010, -0.009)	
Intercept	1.584	0.016	96.530	0.000	(1.551, 1.616)	
Weight < 50 kg (n=317)						0.444 (0.000)
Age	-0.010 (-0.666)	0.001	-15.847	0.000	(-0.011, -0.009)	
Intercept	1.538	0.035	44.394	0.000	(1.469, 1.606)	
50 ≤ Weight < 57.5 kg (n=850)						0.360 (0.000)
Age	-0.010 (-0.600)	0.000	-21.857	0.000	(-0.011, -0.009)	
Intercept	1.565	0.024	64.254	0.000	(1.517, 1.613)	
57.5 ≤ Weight < 65 kg (n=570)						0.347 (0.000)
Age	-0.010 (-0.589)	0.001	-17.360	0.000	(-0.011, -0.009)	
Intercept	1.623	0.031	51.600	0.000	(1.561, 1.685)	
Weight ≥ 65 kg (n=369)						0.328 (0.000)
Age	-0.010 (-0.573)	0.001	-13.888	0.000	(-0.011, -0.009)	
Intercept	1.666	0.041	40.382	0.000	(1.585, 1.748)	

Figure: Overweight vs. Underweight Group

