

경영데이터분석기초

- SPSS, Excel을 활용한 통계분석 -

유 진 호

jhyoo@smu.ac.kr

Measure(측도/척도)

Measurement

- ✓ 미리 결정된 규칙에 의하여 연구대상의 특징을 수치나 기호로 할당하는 과정

*제록없음2 [데이터집합1] - IBM SPSS Statistics Data Editor

파일(F) 편집(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
1	age	숫자	12	1		없음	없음	12	오른쪽	척도(S)	입력
2	job	문자	13	0		없음	없음	13	왼쪽	명목(N)	입력
3	marital	문자	8	0		없음	없음	8	왼쪽	명목(N)	입력
4	education	문자	9	0		없음	없음	9	왼쪽	명목(N)	입력
5	default	문자	3	0		없음	없음	3	왼쪽	명목(N)	입력
6	balance	숫자	12	0		없음	없음	12	오른쪽	척도(S)	입력
7	housing	문자	3	0		없음	없음	3	왼쪽	명목(N)	입력
8	loan	문자	3	0		없음	없음	3	왼쪽	명목(N)	입력
9	contact	문자	9	0		없음	없음	9	왼쪽	명목(N)	입력
10	day	숫자	12	0		없음	없음	12	오른쪽	척도(S)	입력
11	month	문자	3	0		없음	없음	3	왼쪽	명목(N)	입력
12	duration	숫자	12	0		없음	없음	12	오른쪽	척도(S)	입력
13	campaign	숫자	12	0		없음	없음	12	오른쪽	척도(S)	입력
14	pdays	숫자	12	0		없음	없음	12	오른쪽	척도(S)	입력
15	previous	숫자	12	0		없음	없음	12	오른쪽	척도(S)	입력
16	poutcome	문자	7	0		없음	없음	7	왼쪽	명목(N)	입력
17	y	문자	3	0		없음	없음	3	왼쪽	명목(N)	입력

Data Type 정하기,
척도 정하기

- 연속형(수치형)
 - Scale
 - Numeric
- 범주형(이산형/명목형)
 - Nominal
 - Categorical
- 순서형(Ordinal)

◆ 척도에 따른 분석방법 비교

종류	정의	예	구분
명목척도 (nominal)	측정대상(응답자)을 단지 분류할 목적으로 숫자를 부여한 척도	성별, 운동선수의 등번호 등	범주형 자료
서열척도 (ordinal)	측정대상(응답자)간에 순서/서열적 의미(높고 낮음)가 있는 척도	학력, 선호도 등	
등간척도 (interval)	측정대상(응답자)간에 순서 뿐만 아니라 숫자의 간격이 동일하여 양적인 정도를 알 수 있는 척도	온도, 시간 등	연속형 자료
비율척도 (ratio)	등간척도와 유사하나, 0의 의미가 '실제로 없다'고 말할 수 있는 척도	무게, 길이, 소득 등	

X_1 age
 X_2 balance
 X_3 balance
 X_4 default
 \dots
 X_{16}
 Y

$A \rightarrow B$
 $\sigma_1 \rightarrow \sigma_2$
 $\sigma_1 \rightarrow \sigma_2$

\Rightarrow 범주형이 type한 것임.

척도와 분석간의 관계		독립변수 X	
		범주형 자료	연속형 자료
종속변수 Y	범주형 자료	교차분석 (χ^2 검정)	<u>로지스틱 회귀분석</u> 판별분석
	연속형 자료	<u>ANOVA(분산분석)</u>	회귀분석

분석 -> 기술통계

			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)								
			시뮬레이션...								
			품질 관리(Q)								
			ROC 곡선(V)...								
	age	job									
4494	28.0	technician									
4495	26.0	technician									
4496	48.0	management									
4497	30.0	blue-collar									
4498	31.0	entrepreneur									
4499	31.0	management									
4500	45.0	blue-collar									
4501	38.0	admin.									
4502	34.0	management									
4503	42.0	services									
4504	60.0	self-employed									
4505	42.0	blue-collar									
4506	32.0	admin.									
4507	42.0	unemployed									
4508	33.0	services									
4509	42.0	admin.									
4510	51.0	technician									
4511	36.0	technician									
4512	46.0	blue-collar									
4513	40.0	blue-collar									
4514	49.0	blue-collar									
4515	38.0	blue-collar									
4516	32.0	services	single	secondary	no	473	yes	no	cellular	7 jul	624

Data 탐색하기(빈도분석)

.평균
 .표준편차
 .첨도/왜도
 .분포 모양(흩어진 정도)
 .히스토그램
 .Box-Plot 등

이산형 변수와 타겟변수간
 연관성 분석
 .crosstabs(교차분석)

연속형 변수와 타겟변수간
 연관성 분석
 (타겟변수에 영향을 줄 수
 있는 연속형 변수 찾기)
 .Explore(데이터탐색)

	age	job
1	30	unemployed
2	33	services
3	35	management
4	30	management
5	59	blue-collar
6	35	management
7	36	self-employed
8	39	technician
9	41	entrepreneur
10	43	services
11	39	services
12	43	admin.
13	36	technician
14	20	student
15	31	blue-collar
16	40	management
17	56	technician
18	37	admin.
19	25	blue-collar
20	31	services
21	38	management
22	42	management
23	44	services

	income	housing	loan	credit
1	1787	no	no	ce
2	4789	yes	yes	ce
3	1350	yes	no	ce
4	1476	yes	yes	un
5	0	yes	no	un
6	747	no	no	ce
7	307	yes	no	ce
8	147	yes	no	ce
9	221	yes	no	un
10	-88	yes	yes	ce
11	9374	yes	no	un
12	264	yes	no	ce
13	1109	no	no	ce
14	502	no	no	ce
15	360	yes	yes	ce
16	194	no	yes	ce
17	4073	no	no	ce
18	2317	yes	no	ce
19	-221	yes	no	un
20	132	no	no	ce
21	0	yes	no	ce
22	16	no	no	ce
23	106	no	no	un

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(I) ▶

복합 표본(L) ▶

시뮬레이션...

품질 관리(Q) ▶

ROC 곡선(V)...

빈도분석(F)...

기술통계(D)...

데이터 탐색(E)...

교차분석(C)...

비율(R)...

P-P 도표(P)...

Q-Q 도표(Q)...

	age	job	marital	education	default	balance	housing	loan	contact	d
1	30	unemployed	married	primary	no	1787	no	no	cellular	
2	33	services	married	secondary	no	4789	yes	yes	cellular	
3	35	management	single	tertiary	no	1250	yes	no	cellular	
4	30	management	married	tertiary	no	2500	no	no	cellular	
5	59	blue-collar	married	secondary	no	2916	no	no	cellular	
6	35	management	single	tertiary	no	1875	no	no	cellular	
7	36	self-employed	married	tertiary	no	2635	no	no	cellular	
8	39	technician	married	secondary	no	1861	no	no	cellular	
9	41	entrepreneur	married	tertiary	no	891	no	no	cellular	
10	43	services	married	primary	no	4176	yes	yes	cellular	
11	39	services	married	secondary	no	3100	yes	yes	cellular	
12	43	admin.	married	secondary	no	2645	yes	yes	cellular	
13	36	technician	married	tertiary	no	1861	no	no	cellular	
14	20	student	single	secondary	no	2615	no	no	cellular	
15	31	blue-collar	married	secondary	no	1136	no	no	cellular	
16	40	management	married	tertiary	no	1936	no	yes	cellular	
17	56	technician	married	secondary	no	4073	no	no	cellular	
18	37	technician	married	tertiary	no	2594	yes	yes	cellular	

기술통계

변수(V):

age
balance
duration
campaign
pdays
previous

☐ 표준화 값을 변수로 저장(Z)

확인 불여넣기(P) 재설정(R) 취소 도움말

옵션(O)...
붓스트랩(B)...

tertiary	no	1350	yes	no	cellular	16	apr	185
tertiary	no					3	jun	199
secondary	no							226
tertiary	no							141
tertiary	no							341
secondary	no							151
tertiary	no							57
primary	no							313
secondary	no							273
secondary	no							113
tertiary	no							328
secondary	no							261
tertiary	no							89
secondary	no	194	no	yes	cellular			189
secondary	no	4073	no	no	cellular			239
tertiary	no	2317	yes	no	cellular			114
primary	no	-221	yes	no	unknown			250
secondary	no	132	no	no	cellular			148
unknown	no	0	yes	no	cellular			96
tertiary	no	16	no	no	cellular			140
secondary	no	106	no	no	unknown	12	jun	109

기술통계

변수(V):

age
balance
duration
campaign
pdays
previous

표준화 값을 변수로 저장(Z)

확인 불여넣기(P) 재설정(R) 취소 도움말

옵션(O)...
붓스트랩(B)...

기술통계: 옵션

☒ 평균(M) ☐ 합계(S)

산포도

☒ 표준편차(T) ☒ 최소값(N)
☐ 분산(V) ☒ 최대값(X)
☐ 범위(R) ☐ 평균의 표준오차(E)

분포

☐ 첨도(K) ☐ 왜도(W)

출력 순서

☒ 변수목록(B)
☐ 문자순(A)
☐ 평균값 오름차순(C)
☐ 평균값 내림차순(D)

계속 취소 도움말

분석 -> Explore(데이터 탐색)

즐거워요

age Stem-and-Leaf Plot

 $11 * 6 =$

Frequency	Stem &	Leaf
4.00	1 .	&
63.00	2 .	123344&
415.00	2 .	5555666666677777777788888888899999999
990.00	3 .	000000000000011111111111111122222222222222223333333333333333444444444444444444
818.00	3 .	555555555555556666666666666666667777777777777788888888888888999999999999
638.00	4 .	00000000000001111111111122222222222233333333334444444444
565.00	4 .	5555555555666666666666777777777788888888889999999999
443.00	5 .	000000000111111122222222333333333444444
411.00	5 .	555555556666666777777788888888999999
85.00	6 .	00001234
28.00	6 .	569&
23.00	7 .	013&
38.00	Extremes (>=74)	

Stem width: 10
Each leaf: 11 case(s)

11개 보다 작으면 &와 숫자로 표기,
11보다 크면 숫자가 두 개 이상

$$3+7+9+20+24=63$$

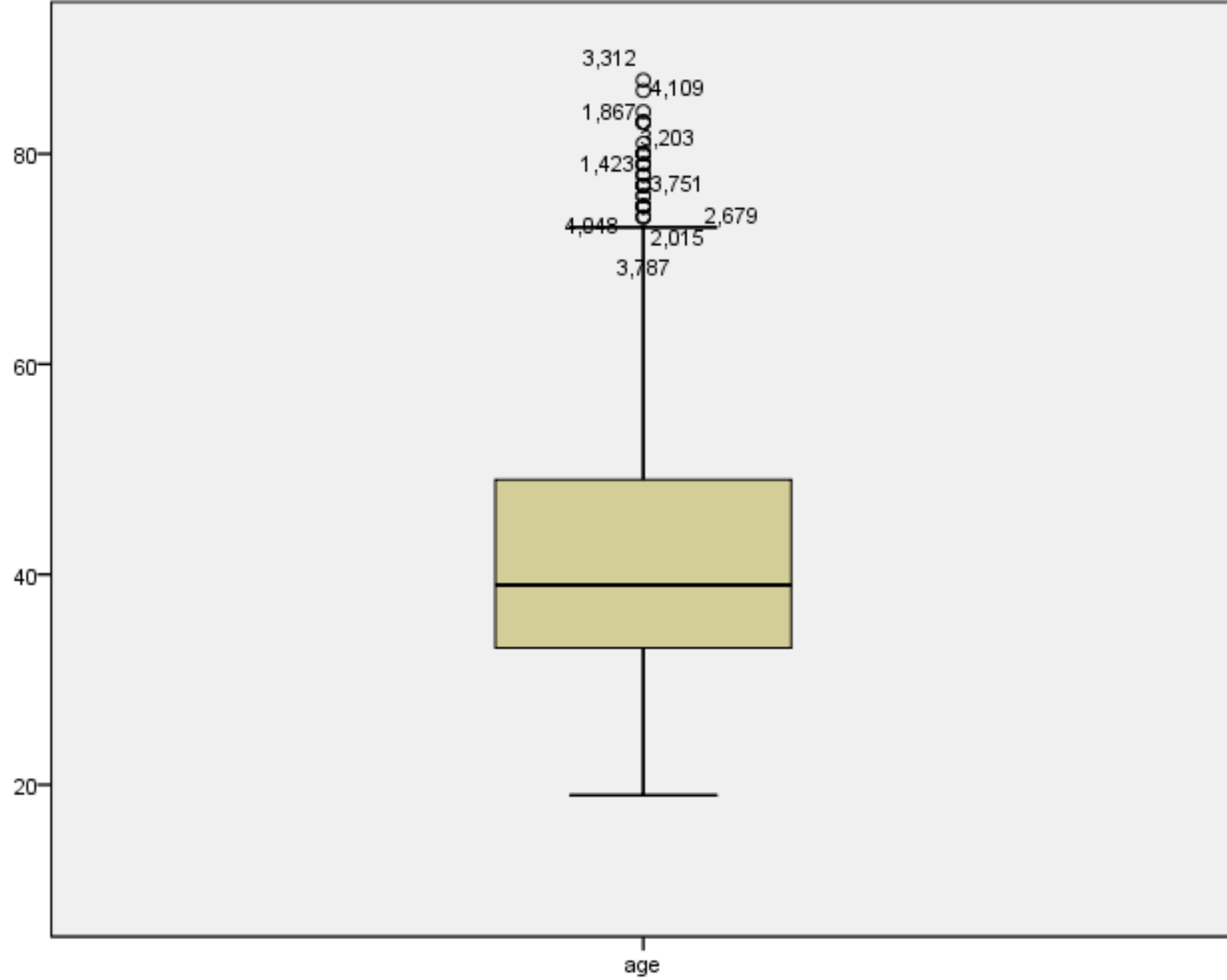
age					
		빈도	퍼센트	유효 퍼센트	누적퍼센트
유효	19	4	.1	.1	.1
	20	3	.1	.1	.2
	21	7	.2	.2	.3
	22	9	.2	.2	.5
	23	20	.4	.4	1.0
	24	24	.5	.5	1.5
	25	44	1.0	1.0	2.5
	26	77	1.7	1.7	4.2
	27	94	2.1	2.1	6.2
	28	103	2.3	2.3	8.5
	29	97	2.1	2.1	10.7
	30	150	3.3	3.3	14.0
	31	199	4.4	4.4	18.4
	32	224	5.0	5.0	23.3
	33	186	4.1	4.1	27.4
	34	231	5.1	5.1	32.6
	35	180	4.0	4.0	36.5
	36	188	4.2	4.2	40.7
	37	161	3.6	3.6	44.3
	38	159	3.5	3.5	47.8
	39	130	2.9	2.9	50.7
	40	142	3.1	3.1	53.8
	41	135	3.0	3.0	56.8
	42	141	3.1	3.1	59.9
	43	115	2.5	2.5	62.4
	44	105	2.3	2.3	64.8
	45	112	2.5	2.5	67.2
	46	119	2.6	2.6	69.9
	47	108	2.4	2.4	72.3
	48	114	2.5	2.5	74.8
	49	112	2.5	2.5	77.3
	50	101	2.2	2.2	79.5

$$7+6+4+6=23$$

38

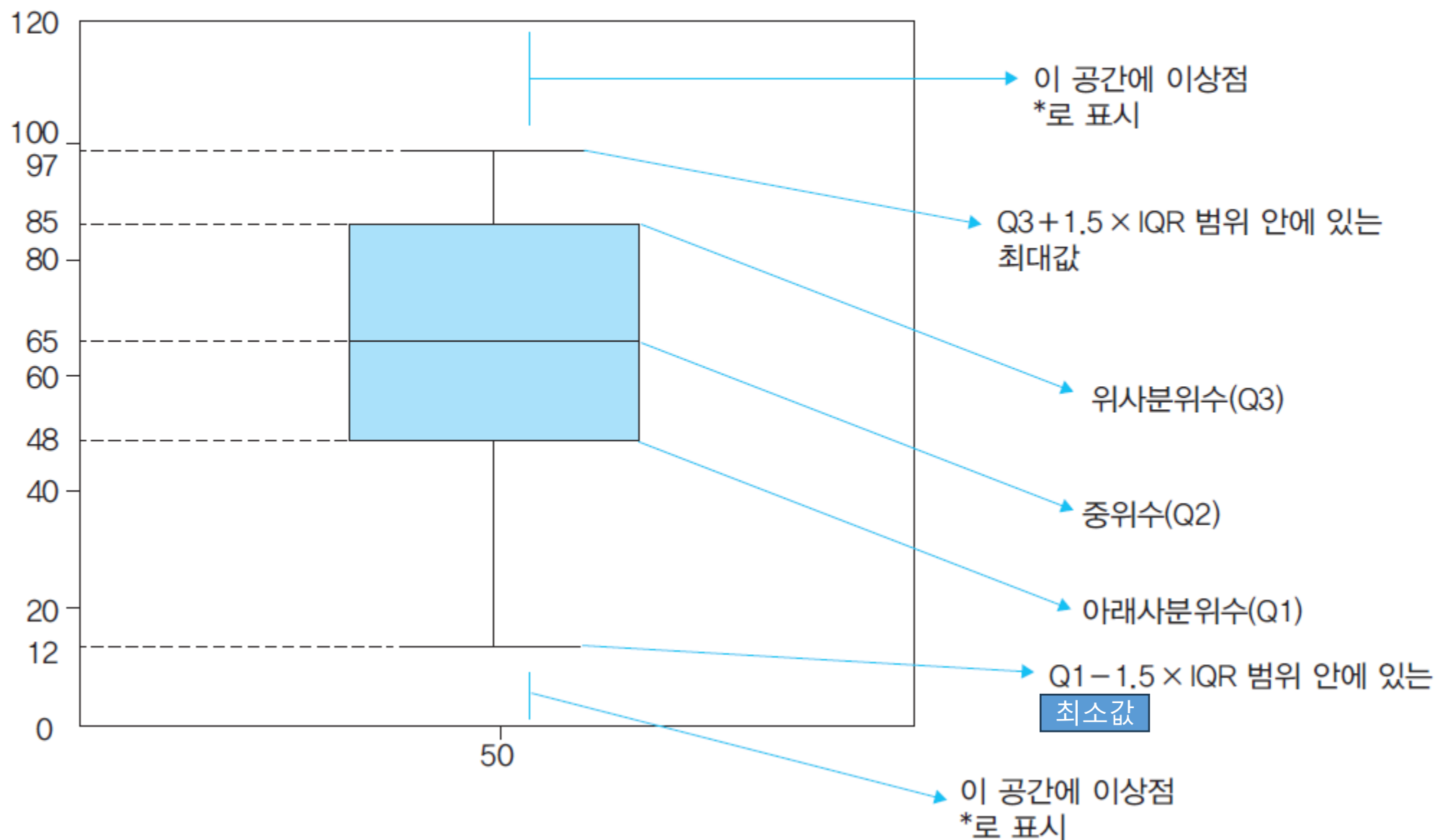
60	47	1.0	1.0	97.2
61	16	.4	.4	97.5
62	7	.2	.2	97.7
63	8	.2	.2	97.9
64	7	.2	.2	98.0
65	6	.1	.1	98.2
66	9	.2	.2	98.4
67	5	.1	.1	98.5
68	2	.0	.0	98.5
69	6	.1	.1	98.7
70	7	.2	.2	98.8
71	6	.1	.1	98.9
72	4	.1	.1	99.0
73	6	.1	.1	99.2
74	3	.1	.1	99.2
75	6	.1	.1	99.4
76	2	.0	.0	99.4
77	6	.1	.1	99.5
78	3	.1	.1	99.6
79	4	.1	.1	99.7
80	6	.1	.1	99.8
81	1	.0	.0	99.8
83	4	.1	.1	99.9
84	1	.0	.0	100.0
86	1	.0	.0	100.0
87	1	.0	.0	100.0
합계	4521	100.0	100.0	

BOX PLOT

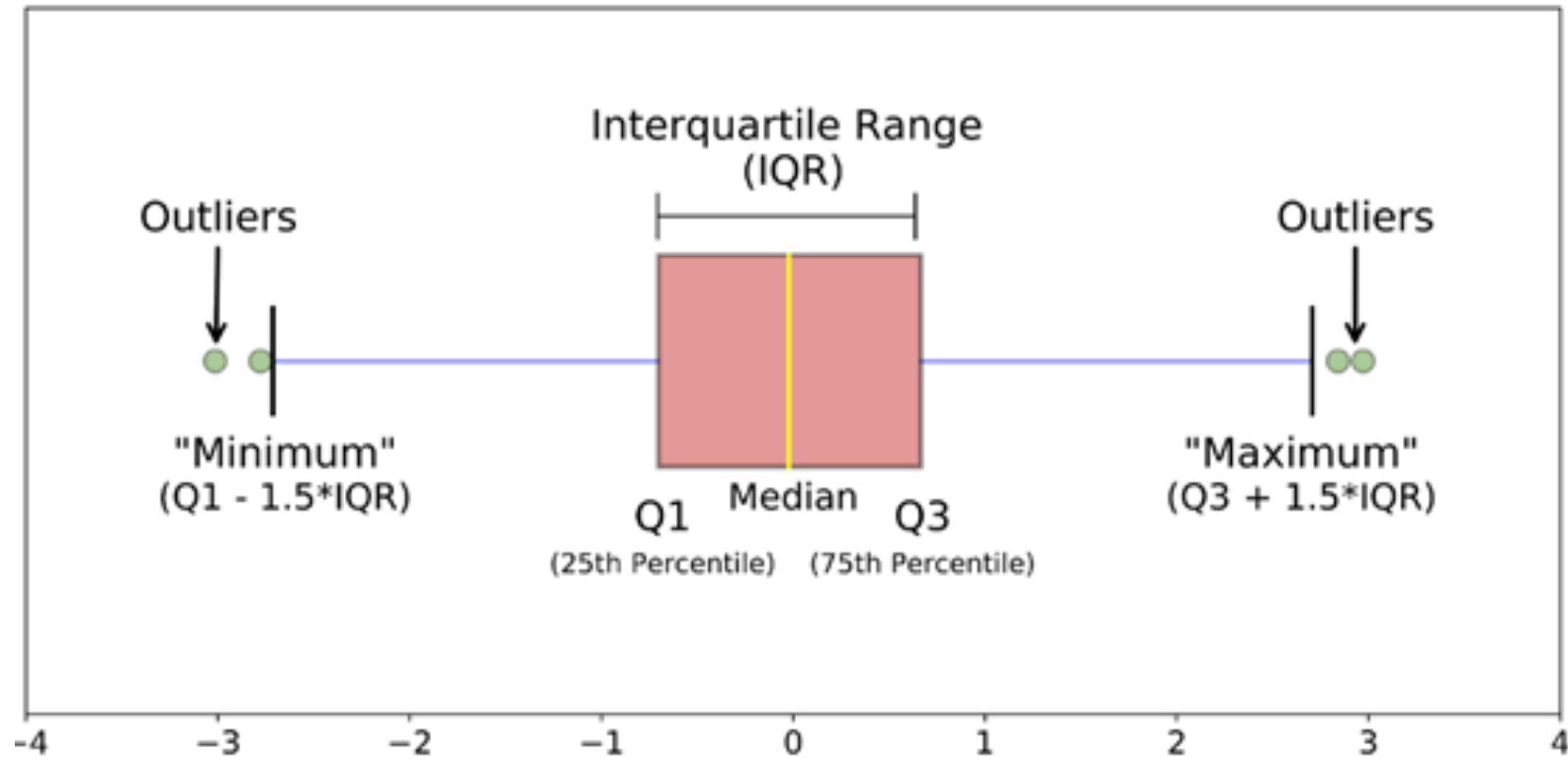


BoxPlot 그림 이해하기

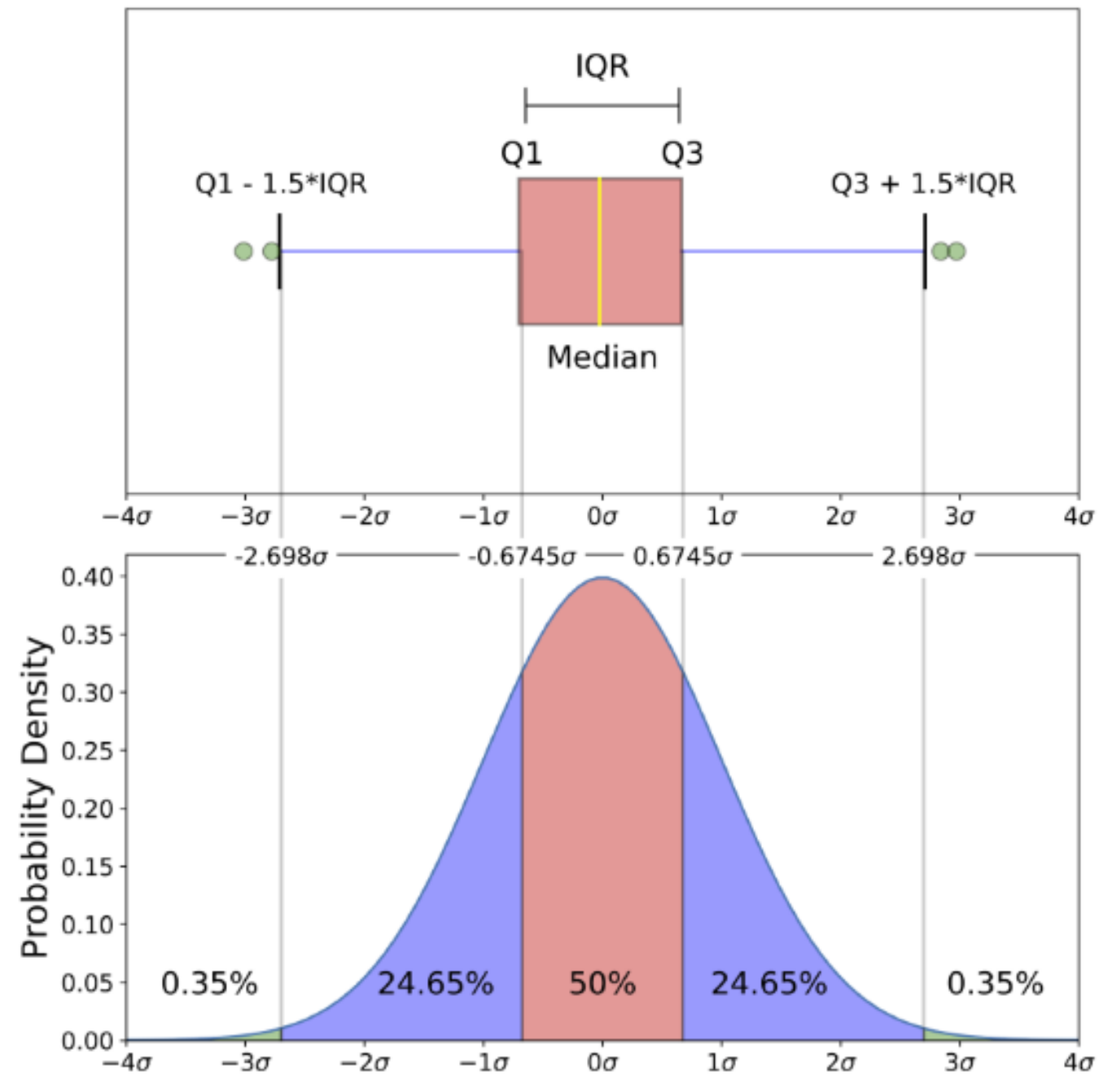
$$IQR = Q_3 - Q_1$$



BOX PLOT



$$IQR = Q_3 - Q_1$$



99.3%
0.7%

BOX PLOT

- 분석 -> 기술통계량 -> Explore(데이터탐색) 선택

bank.sav [데이터집합1] - IBM SPSS Statistics Data Editor

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

age job marital education default balance housing loan contact da

1						1787	no	no	cellular	
2						4789	yes	yes	cellular	
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16	40	management	married	tertiary	no					
17	56	technician	married	secondary	no					
18	37	admin	single	tertiary	no					

데이터 탐색

종속변수(D): age

요인(F):

케이스 설명 기준변수(C):

표시: ☒ 모두(B) ☐ 통계량(A) ☐ 도표(L)

통계량(S)...
도표(T)...
옵션(O)...
붓스트랩(B)...

데이터 탐색: 도표

상자도표: ☒ 요인수준들과 함께(E) ☐ 종속변수들과 함께(D) ☐ 지정없음(N)

기술통계: ☒ 줄기와 앞그림(S) ☒ 히스토그램(H)

☐ 검정과 함께 정규성도표(O)

Levene 검정이 있는 평균-산포: ☒ 없음(E) ☐ 제공값 추정(P) ☐ 변환(T) 제공값(W): 자연로그 ☐ 변환하지 않음(U)

계속 취소 도움말

2. box-plot

1. 정규성 검정

Q-Q plot

$k-h$, $h-w$

정규성 검증

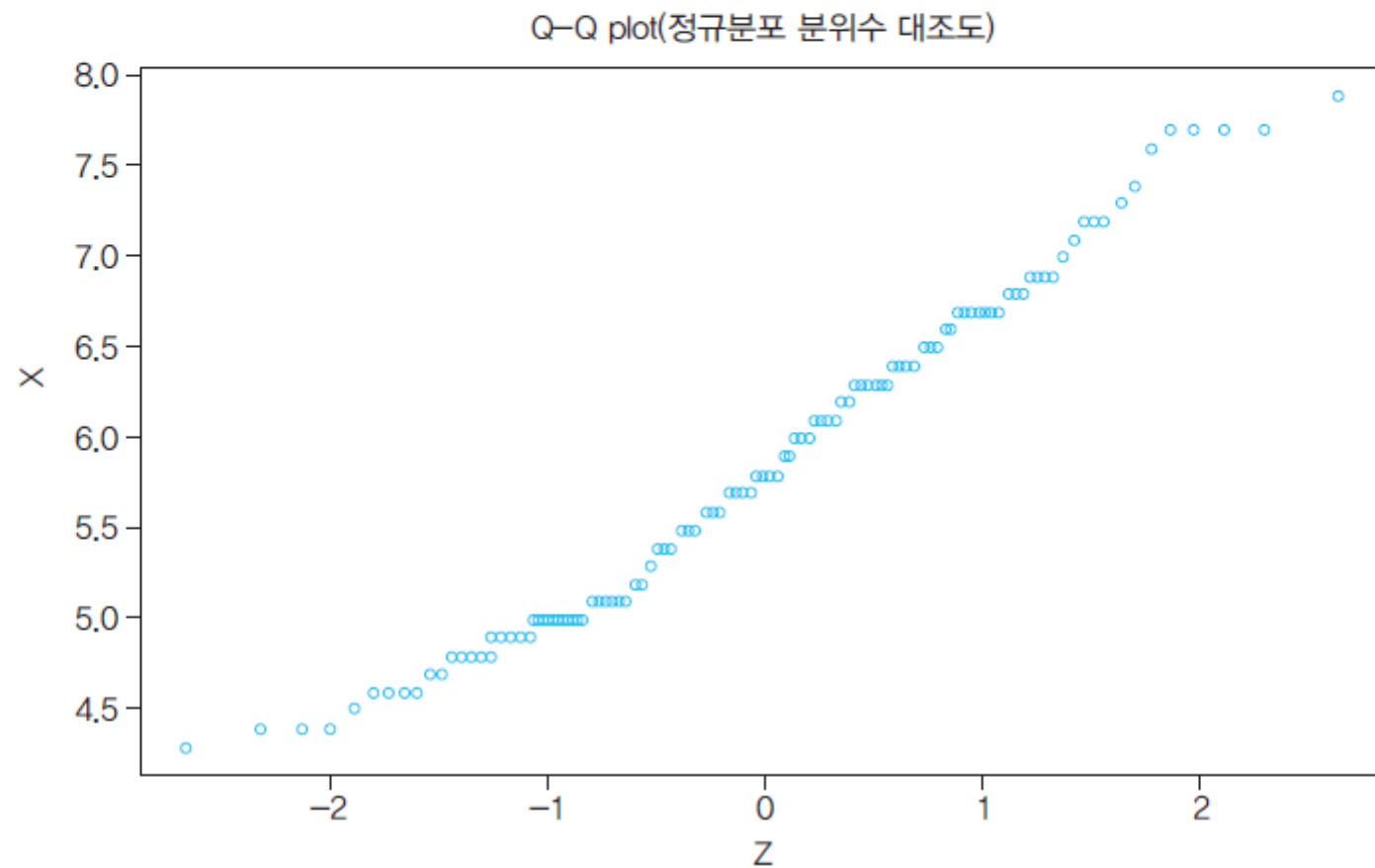
◆ 통계 - 정규성 검토 : Q-Q plot (Quantile-Quantile Plot)

통계 추론은 대부분이 모집단이 정규분포를 따른다는 가정에 진행된다.

이 이유로 통계적 추론의 방법을 적용하기 전 데이터에 대한 정규모집단의 가정에 대한 검토가 필요하다. 정규분포 가정을 하는 방법 중 Q-Qplot(정규분포 분위수 대조도)를 그리는 방법이다.

◆ Q-Q plot이란

qqplot은 분위수대조도로 불리며 정규모집단 가정을 하는 방법 중 하나이며 수집 데이터를 표준정규분포의 분위수와 비교하여 그리는 그래프이다. 이는 간단하게 데이터의 정규성 가정에 대한 검토를 가능하게 해 준다. 아래의 그림이 qqplot의 일종으로 모집단이 정규성을 따른다면 아래와 같이 직선의 형태로 그려지게 된다.



위 사분위수에 대한 개념이 분위수 사용의 예라고 볼 수 있다.

qqplot에서도 마찬가지로 위와 같은 분위수 개념이 적용된다.

이는 표준정규분포의 분위수와 이에 대응하는 분포의 분위수를 x , y 좌표평면에 plotting 하는 개념이다.

확률 표본($X_1, X_2, X_3, X_4 \dots X_{100}$)이 평균이 m , 표준편차가 s 인 정규모집단을 따르는 모집단에서 추출된 표본인지에 대한 검토를 예로 들어보자.

표준정규분포의 분위수는 $1 \sim n$ (표본의 수)의 백분율 값만큼의 표준정규분포 X 좌표 value가 된다.

예제의 분위수는 $Z_{0.01}, Z_{0.02}, \dots Z_{1.00}$ 의 값이 된다.

표준정규분포의 분위수 값과 표본 데이터를 크기순으로 정렬($X_1 < X_2 < X_3 \dots X_n$) 한 값을 가지고 plotting 하게 되면 qqplot을 얻을 수 있게 된다.

가장 일반적으로 qqplot 내 점들이 기울기가 45도인 직선에 점들이 밀집되어 있어야 정규분포를 가정할 수 있다.

age

	빈도	퍼센트	유효 퍼센트	누적퍼센트
유효	19	4	.1	.1
	20	3	.1	.2
	21	7	.2	.3
	22	9	.2	.5
	23	20	.4	1.0
	24	24	.5	1.5
	25	44	1.0	2.5
	26	77	1.7	4.2
	27	94	2.1	6.2
	28	103	2.3	8.5
	29	97	2.1	10.7
	30	150	3.3	14.0
	31	199	4.4	18.4
	32	224	5.0	23.3
	33	186	4.1	27.4
	34	231	5.1	32.6
	35	180	4.0	36.5
	36	188	4.2	40.7
	37	161	3.6	44.3
	38	159	3.5	47.8
	39	130	2.9	50.7
	40	142	3.1	53.8
	41	135	3.0	56.8
	42	141	3.1	59.9
	43	115	2.5	62.4
	44	105	2.3	64.8
	45	112	2.5	67.2
	46	119	2.6	69.9
	47	108	2.4	72.3
	48	114	2.5	74.8
	49	112	2.5	77.3
	50	101	2.2	79.5

10% 지점

20% 지점

30% 지점

40% 지점

50% 지점

표준정규분포표 [편집]

z보다 작은 값을 가질 확률에 대한 누적 표준정규분포표는 다음과 같다.^[3]

↓ 값의 반대

-2.31, -2.05, -1.88, -1.75, 0, 0.26, 0.53, 0.85, 1.29

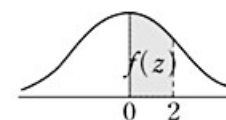
↓ 0.0000

z	- 0.00	- 0.01	- 0.02	- 0.03	- 0.04	- 0.05	- 0.06	- 0.07	- 0.08	- 0.09
-4.0	0.00003	0.00003	0.00003	0.00003	0.00003	0.00003	0.00002	0.00002	0.00002	0.00002
-3.9	0.00005	0.00005	0.00004	0.00004	0.00004	0.00004	0.00004	0.00004	0.00003	0.00003
-3.8	0.00007	0.00007	0.00007	0.00006	0.00006	0.00006	0.00006	0.00005	0.00005	0.00005
-3.7	0.00011	0.00010	0.00010	0.00010	0.00009	0.00009	0.00008	0.00008	0.00008	0.00008
-3.6	0.00016	0.00015	0.00015	0.00014	0.00014	0.00013	0.00013	0.00012	0.00012	0.00011
-3.5	0.00023	0.00022	0.00022	0.00021	0.00020	0.00019	0.00019	0.00018	0.00017	0.00017
-3.4	0.00034	0.00032	0.00031	0.00030	0.00029	0.00028	0.00027	0.00026	0.00025	0.00024
-3.3	0.00048	0.00047	0.00045	0.00043	0.00042	0.00040	0.00039	0.00038	0.00036	0.00035
-3.2	0.00069	0.00066	0.00064	0.00062	0.00060	0.00058	0.00056	0.00054	0.00052	0.00050
-3.1	0.00097	0.00094	0.00090	0.00087	0.00084	0.00082	0.00079	0.00076	0.00074	0.00071
-3.0	0.00135	0.00131	0.00126	0.00122	0.00118	0.00114	0.00111	0.00107	0.00104	0.00100
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139
-2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
-2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
-2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.3	0.01072	<u>0.01044</u>	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2.0	0.02275	0.02222	0.02169	0.02118	0.02068	<u>0.02018</u>	0.01970	0.01923	0.01876	0.01831
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	<u>0.03005</u>	0.02938
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	<u>0.04006</u>	0.03920	0.03836	0.03754	0.03673
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551
-1.5	0.06681	0.06552	0.06426	0.06301	0.06178	0.06057	0.05938	0.05821	0.05705	0.05592

z	+ 0.00	+ 0.01	+ 0.02	+ 0.03	+ 0.04	+ 0.05	+ 0.06	+ 0.07	+ 0.08	+ 0.09
0.0	<u>0.50000</u>	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	<u>0.60257</u>	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	<u>0.70194</u>	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	<u>0.80234</u>	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	<u>0.90147</u>
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91308	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169

정규분포표

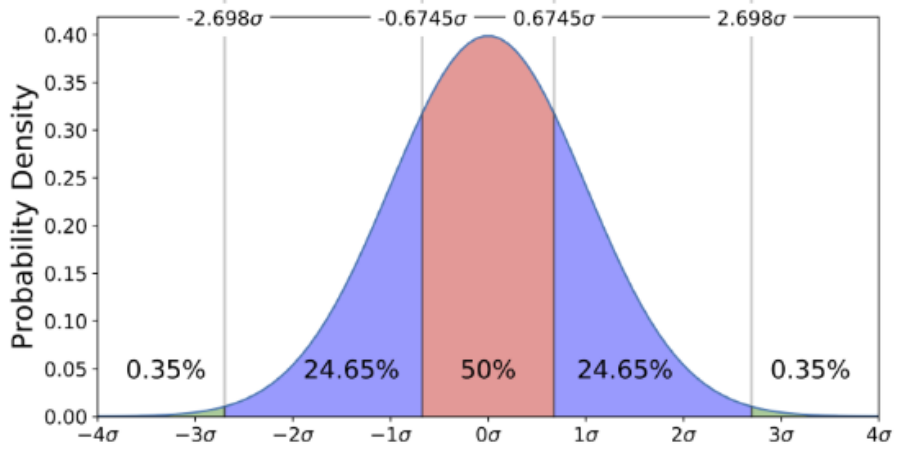
$$f(z) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}z^2} dz$$



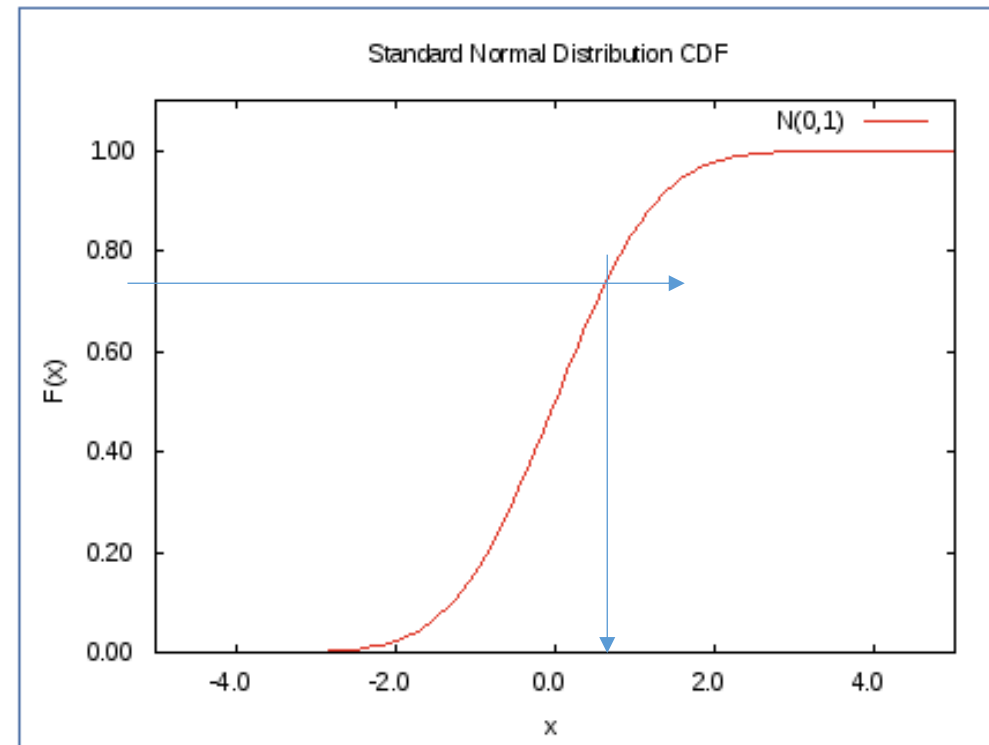
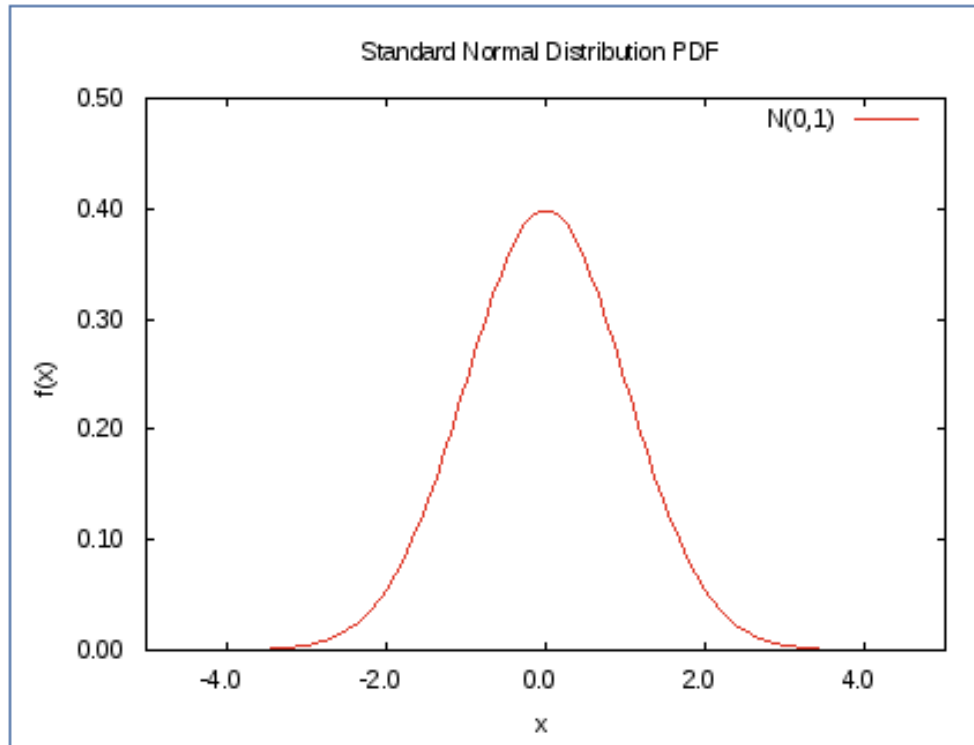
0.6745

25% 지점
누적: 75% 지점

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441

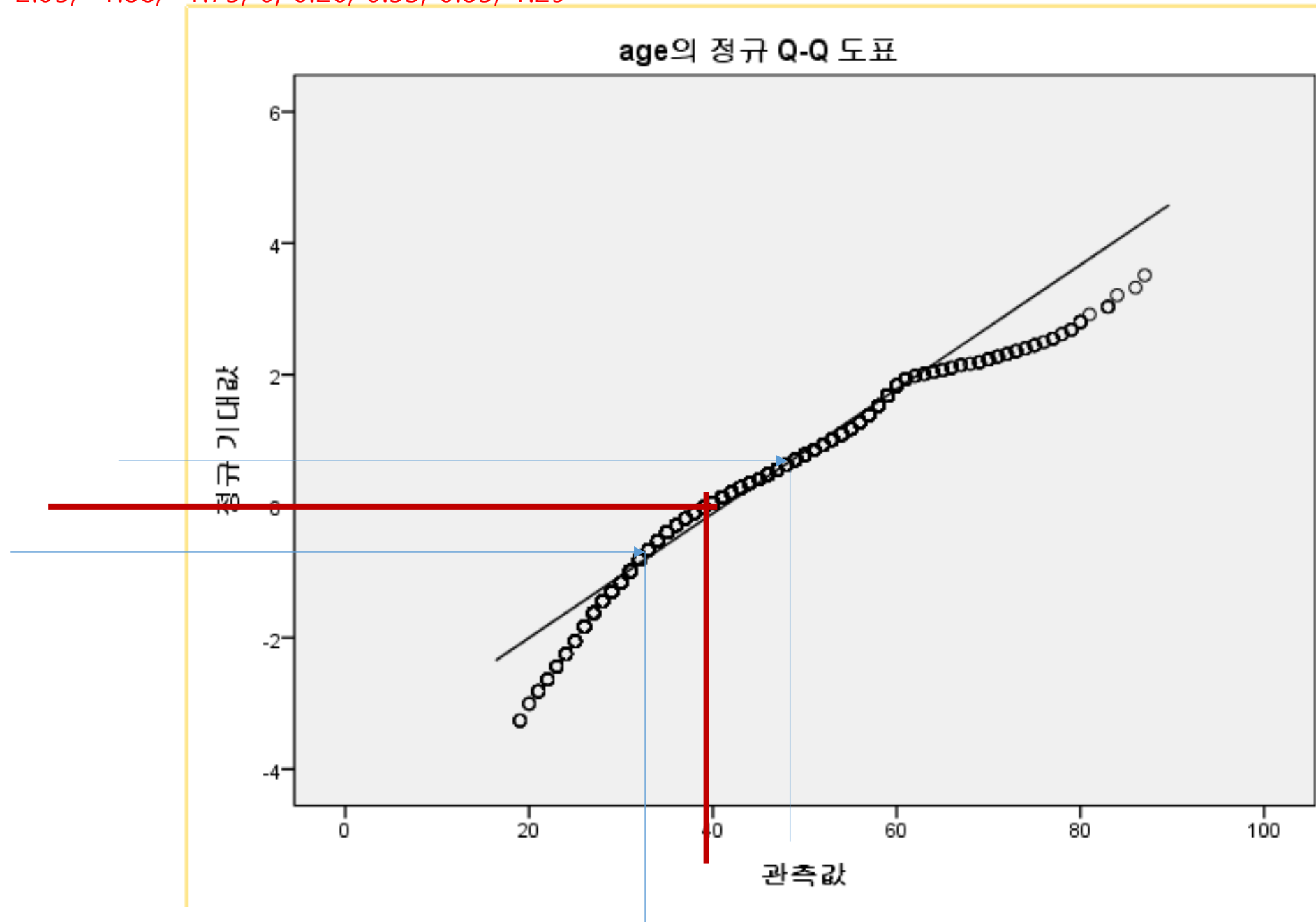


$$F(x)=0.75$$



$$X=0.675..$$

-2.31, -2.05, -1.88, -1.75, 0, 0.26, 0.53, 0.85, 1.29

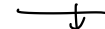


◆ 정규성 검토의 다른 방법

정규성 검토에는 qqplot을 그려 검토하는 방법 말고도 여러 가지 방법이 존재한다. 가장 간단하게 확인하려 한다면 히스토그램을 그려 그 자료의 분포를 살펴는 방법이 있다. 표본의 분포를 histogram이나 density plot을 그려 정규분포와 유사한가 살펴는 방법이다. shapiro-test 방법도 있다.

⊕ 유의수준 = 0.05

유의수준이 0.05보다 작으면 귀무가설 기각.



이때 귀무가설 H_0 : 정규분포를 따른다

⇒ 정규분포를 따르지 않는다.

정규성 검토

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	통계량	자유도	유의확률	통계량	자유도	유의확률
age	.096	4521	.000	.960	4521	.000

a. Lilliefors 유의확률 수정

집단별로 age를 분석하자.

예) default(채무불이행) 여부별 age 분포는?

차이 X

숙제)

y(정기예금 상품 가입여부) 집단별 duration 분포는?

. 데이터탐색 분석하고 결과파일을 *.xlsx 파일로 저장해서, 해석달기

하이오

default는 yes이거임,
default는 no이거임이
2개씩 차이가 있음