

# 기술통계량

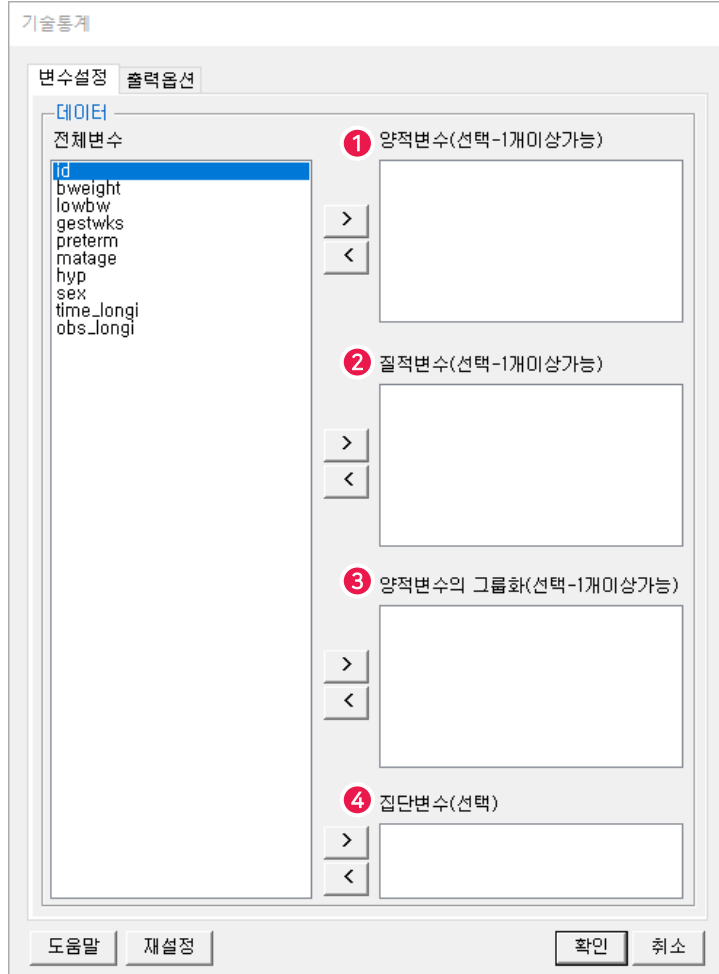
기술 통계량은 표본 자체의 속성을 파악하기 위하여 데이터 요약을 제공하는 통계분석 모듈입니다. 기술통계는 데이터의 속성을 특정한 통계량을 사용해서 그 요약을 보여줍니다. 기술통계는 매우 간단한 통계량으로 엄청난 양의 데이터가 갖는 속성을 합리적인 방법으로 간결하게 요약해줌으로써, 연구자가 데이터의 속성을 쉽게 이해할 수 있도록 도와주기에 매우 중요한 분석이라고 할 수 있습니다.

## 메뉴 호출하기

- Rex > 기술통계량



• 변수설정 탭



메뉴 요소	설명
① 양적변수	연속형 변수에 해당하는 변수를 전체변수로부터 선택합니다. 한 개 이상의 변수가 선택될 수 있으며, 질적변수와 중복되어 선택될 수 없습니다.
② 질적변수	범주형 변수에 해당하는 변수를 전체변수로부터 선택합니다. 한 개 이상의 변수가 선택될 수 있으며, 양적변수와 중복되어 선택될 수 없습니다.
③ 양적변수의 그룹화	연속형 변수를 특정 구간으로 구분하여 그룹화할 때 사용됩니다. 연속형 변수를 전체변수로부터 한 개 이상 선택하며, 양적변수와 중복되어 선택될 수 없습니다.
④ 집단변수	그룹별로 나누어 연속형 또는 범주형 변수의 기술통계량을 구하고자 할 때 사용됩니다. 그룹으로 지정하고자 하는 한 개의 범주형 변수를 전체변수로부터 선택합니다. 양적변수, 질적변수와 중복되어 선택될 수 없습니다.

출력옵션 탭

메뉴 요소	설명
① 통계량 > 양적변수	<p>변수설정 탭에서 양적변수가 1개 이상 선택된 경우 활성화되며, 선택된 통계량이 결과로 출력됩니다. 전체선택/해제 버튼을 이용하여 모든 옵션을 선택 혹은 해제할 수 있습니다.</p> <ul style="list-style-type: none"> <li>• 평균 : 각 변수의 관측값의 평균을 계산합니다.</li> <li>• 중앙값 : 각 변수의 관측값의 중앙값을 계산합니다.</li> <li>• 최소값 : 각 변수의 관측값의 최소값을 계산합니다.</li> <li>• 최대값 : 각 변수의 관측값의 최대값을 계산합니다.</li> <li>• 제1사분위수 : 각 변수의 관측값의 25% 백분위수를 계산합니다.</li> <li>• 제3사분위수 : 각 변수의 관측값의 75% 백분위수를 계산합니다.</li> <li>• 범위 : 각 변수의 최소값과 최대값 사이의 길이를 계산합니다.</li> <li>• 합계 : 각 변수의 관측값의 합을 계산합니다.</li> <li>• 분산 : 각 변수의 관측값의 표본분산을 계산합니다.</li> <li>• 표준편차 : 각 변수의 관측값의 표본표준편차를 계산합니다.</li> <li>• 왜도 : 각 변수의 관측값의 표본 왜도를 계산합니다.</li> <li>• 첨도 : 각 변수의 관측값의 표본 첨도를 계산합니다.</li> <li>• 변동계수(CV) : 각 변수의 변동계수(표준편차÷평균)를 계산합니다.</li> </ul>
② 통계량 > 질적변수	<p>변수설정 탭에서 질적변수가 1개 이상 선택된 경우 활성화되며, 선택된 통계량이 결과로 출력됩니다. 전체선택/해제 버튼을 이용하여 모든 옵션을 선택 혹은 해제할 수 있습니다.</p> <ul style="list-style-type: none"> <li>• 빈도 : 각 변수의 수준별 관측값의 빈도를 계산합니다.</li> <li>• 누적빈도 : 각 변수의 수준별 관측값의 누적빈도를 계산합니다.</li> <li>• 상대빈도 : 각 변수의 수준별 관측값의 상대빈도를 계산합니다.</li> <li>• 누적상대빈도 : 각 변수의 수준별 관측값의 누적상대빈도를 계산합니다.</li> <li>• 퍼센트 : 각 변수의 수준별 퍼센트를 계산합니다.</li> <li>• 누적퍼센트 : 각 변수의 수준별 누적퍼센트를 계산합니다.</li> <li>• 유효퍼센트 : 각 변수의 수준별 유효퍼센트를 계산합니다.</li> <li>• 누적유효퍼센트 : 각 변수의 수준별 누적유효퍼센트를 계산합니다.</li> </ul>

• 출력옵션 탭

메뉴 요소	설명
③ 양적변수의 그룹화 > 계급의 수	<p>변수설정 탭에서 양적변수의 그룹화에 1개 이상의 양적변수가 선택된 경우 활성화됩니다. 선택된 계급의 수에 맞게 양적변수를 그룹화하여, 각 그룹별 관측값의 빈도를 계산합니다. 계급의 수는 관측값의 개수가 <math>n</math>일 때, 다음의 4가지 옵션 중 1개를 선택할 수 있습니다.</p> <ul style="list-style-type: none"> <li>Sturges 공식 (디폴트): 계급의 수를 <math>\lceil \log_2 n \rceil</math> 으로 계산합니다.</li> <li>제곱근방법: 계급의 수를 <math>\lceil \sqrt{n} \rceil</math> 으로 계산합니다.</li> <li>Rice 공식: 계급의 수를 <math>\lceil 2n^{1/3} \rceil</math> 으로 계산합니다.</li> <li>사용자정의: 사용자가 원하는 계급의 수를 양의 정수로 직접 입력하여 결정합니다.</li> </ul>
④ 양적변수의 그룹화 > 계급의 폭	<p>변수설정 탭에서 양적변수의 그룹화에 1개 이상의 연속형 변수가 선택된 경우 활성화됩니다. 선택된 계급의 폭에 맞게 양적변수를 그룹화하여, 각 그룹별 관측값의 빈도를 계산합니다. 계급의 폭은 다음의 2가지 옵션 중 1개를 선택할 수 있습니다.</p> <ul style="list-style-type: none"> <li>자동계산: <math>\left[ \frac{\text{최대값} - \text{최소값}}{\text{계급의 수}} \right]</math> 으로 계급의 폭을 계산합니다.</li> <li>사용자정의: 사용자가 원하는 계급의 폭을 양수로 입력받아 결정합니다.</li> </ul>
⑤ 그래프 > 양적변수	<p>변수설정 탭에서 양적변수가 1개 이상 선택된 경우 활성화되며, 히스토그램, 박스그림, 줄기잎그림, QQ그림 중 선택된 그래프가 결과로 출력됩니다.</p>
⑥ 그래프 > 질적변수	<p>변수설정 탭에서 질적변수가 1개 이상 선택된 경우 활성화되며, 막대그림과 원형표 중 선택된 그래프가 결과로 출력됩니다. 원형표를 선택할 경우, 빈도 또는 퍼센트 옵션이 활성화되며, 선택된 통계량이 그래프 위에 표기됩니다.</p>

• 출력옵션 탭

기술통계

변수설정 출력옵션

통계량

1 양적변수

평균     중앙값

최소값     최대값

제1사분위수     제3사분위수

범위     합계

분산     표준편차

왜도     첨도

변동계수

전체선택/해제

2 질적변수

빈도     누적빈도

상대빈도     누적상대빈도

퍼센트     누적퍼센트

유효퍼센트

누적유효퍼센트

전체선택/해제

양적변수의 그룹화

3 계급의 수

Sturges 공식     제곱근방법

Rice 공식     사용자정의

4 계급의 폭

자동계산

사용자정의

그래프

5 양적변수

히스토그램

박스그림

줄기잎그림

QQ그림

6 질적변수

막대그림

원형표

빈도

퍼센트

7 요약표

요약표 출력

도움말    재설정    확인    취소

메뉴 요소

설명

변수설정 탭에서 선택된 양적변수 또는 질적변수의 기술통계량을 활용하여 요약표를 생성합니다.

- 변수설정 탭에서 집단변수가 선택되지 않은 경우, 전체 개체에 대한 기술통계량을 출력하며, 집단변수가 선택된 경우, 전체 개체 및 집단별 기술통계량을 출력합니다.
- 연속형 변수의 경우 관측값이 1000개 미만인 경우 Shapiro-Wilk test, 1000개 이상인 경우 Kolmogorov-Smirnov test의 정규성 검정 결과에 따라 평균과 표준편차 또는 중앙값과 사분위수 범위가 출력됩니다.
- 범주형 변수의 경우, 빈도와 그룹별 백분율(%)이 출력됩니다.
- 첫 번째 요약표(Summary table)의 경우, 정규성 검정 결과 또는 카이 제곱검정 가정 위반 여부에 따라 적합한 결과를 선택적으로 출력합니다.
- 두 번째 상세 요약표(Summary table in detail)의 경우, 모수적 검정과 비모수적 검정 결과를 순차적으로 출력하며, 연속형 변수의 경우 정규성 검정 결과와 전체 및 그룹별 범위(최소값, 최대값)를 출력합니다.

⑦ 요약표

• 예제

산모의 고혈압 여부(hyp)에 따른 출생체중(bweight), 저체중여부(lowbw), 재태주수(gestwks), 조기 출산(preterm), 산모의 연령(matage), 신생아의 성별(sex)에 대한 요약비교표를 생성하시오 [데이터 : manual\_data.xlsx > birth].

1. [변수설정] > [양적변수] > 'bweight', 'gestwks', 'matage' 지정
2. [변수설정] > [질적변수] > 'lowbw', 'preterm', 'sex' 지정
3. [변수설정] > [집단변수] > 'hyp' 지정
4. [출력옵션] > [요약표] 선택
5. [확인]

• 예제 - 결과창

**Descriptive Statistics**

**Table Report**

Table.

Variable	Category	Statistics	0 (N=428)	1 (N=72)	Total (N=500)
bweight		n	428	72	500
		mean ± Std	3198.90 ± 586.59	2768.21 ± 790.84	3136.88 ± 637.45
		median	3252.00	2980.50	3188.50
		Q1, Q3	2917.00, 3571.00	2414.50, 3220.50	2862.00, 3551.25
		min ~ max	628.00 ~ 4553.00	693.00 ~ 4436.00	628.00 ~ 4553.00
gestwks		n	419	71	490
		mean ± Std	38.93 ± 2.09	37.51 ± 3.08	38.72 ± 2.31
		median	39.35	38.03	39.12
		Q1, Q3	38.15, 40.17	36.97, 39.39	37.94, 40.09
		min ~ max	24.69 ~ 43.16	27.33 ~ 42.20	24.69 ~ 43.16
matage		n	428	72	500
		mean ± Std	34.13 ± 3.81	33.42 ± 4.38	34.03 ± 3.90
		median	34.00	33.00	34.00
		Q1, Q3	32.00, 37.00	31.00, 37.00	31.00, 37.00
		min ~ max	23.00 ~ 43.00	24.00 ~ 43.00	23.00 ~ 43.00
lowbw	0	n(%)	388(90.65%)	52(72.22%)	440(88.00%)
	1	n(%)	40(9.35%)	20(27.78%)	60(12.00%)
	Total	n(%)	428(100.00%)	72(100.00%)	500(100.00%)
preterm	0	n(%)	375(89.50%)	52(73.24%)	427(87.14%)
	1	n(%)	44(10.50%)	19(26.76%)	63(12.86%)
	Total	n(%)	419(100.00%)	71(100.00%)	490(100.00%)
sex	1	n(%)	221(51.64%)	43(59.72%)	264(52.80%)
	2	n(%)	207(48.36%)	29(40.28%)	236(47.20%)
	Total	n(%)	428(100.00%)	72(100.00%)	500(100.00%)

[Table Report]

- 집단변수 hyp에 따른 변수들의 비교 결과를 한 눈에 볼 수 있는 요약표입니다.

• 예제 - 결과창

**Data Structure**

No. of total observations	500
No. of used variable(s)	7

**Variable List**

Category	Variable	N	N.valid	(%.valid)	N.miss	(%.miss)
Continuous variable	bweight	500	500	(100.00%)	0	(0.00%)
Continuous variable	gestwks	500	490	(98.00%)	10	(2.00%)
Continuous variable	matage	500	500	(100.00%)	0	(0.00%)
Categorical variable	lowbw	500	500	(100.00%)	0	(0.00%)
Categorical variable	preterm	500	490	(98.00%)	10	(2.00%)
Categorical variable	sex	500	500	(100.00%)	0	(0.00%)
Group variable	hyp	500	500	(100.00%)	0	(0.00%)

- Warning : 'lowbw, preterm, sex' are numeric but were chosen as qualitative variable. They were coerced into character.

**Results of Descriptive Statistics**

**Continuous Variable**

Var.1	Var.2	hyp : 0	hyp : 1
bweight	N. observed	428	72
	N. missing	0	0
	Mean	3198.90	2768.21
	Median	3252.00	2980.50
	Min	628.00	693.00
	Max	4553.00	4436.00
	Q1	2917.00	2414.50
	Q3	3571.00	3220.50
	Range	3925.00	3743.00
	Sum	1369131.00	199311.00
	SD	586.59	790.84
	Variance	344092.62	625431.46
	Skewness (Pearson)	-0.89	-0.74
	Skewness (Fisher)	-0.89	-0.75
	SE (Skewness)	0.12	0.28
	Kurtosis (Pearson)	2.15	0.32
	Kurtosis (Fisher)	2.19	0.43
	SE (Kurtosis)	0.24	0.56
	CV	0.18	0.29

[Data Structure]

- No. of total observations : 전체 개체 수
- No. of used variable(s) : 분석에 사용된 변수의 수

[Variable List]

- 분석에 사용된 개별 변수들의 유효개수(N.valid), 유효퍼센트(%.valid), 결측수(N.miss), 결측퍼센트(%.miss)가 출력됩니다.

[Results of Descriptive Statistics]

[Continuous Variables]

- [출력옵션] > [통계량] > [양적변수]에서 지정된 통계량이 변수별로 출력됩니다.



• 예제 - 결과창

Normality test

bweight

Variable		Level	Statistic	P-value
bweight	Shapiro-Wilk test	0	0.956	<0.0001
		1	0.945	0.0033
	Kolmogorov-Smirnov test	0	0.071	0.0273
		1	0.125	0.2101

gestwks

Variable		Level	Statistic	P-value
gestwks	Shapiro-Wilk test	0	0.844	<0.0001
		1	0.831	<0.0001
	Kolmogorov-Smirnov test	0	0.111	0.0001
		1	0.190	0.0118

matage

Variable		Level	Statistic	P-value
matage	Shapiro-Wilk test	0	0.981	<0.0001
		1	0.983	0.4195
	Kolmogorov-Smirnov test	0	0.088	0.0028
		1	0.072	0.8498

- It is recommended to use shapiro-wilk test if sample size is less than 2,000, and kolmogorov-smirnov test if sample size is 2,000 or more.

Categorical Variable

lowbw

hyp - 0

Level	Frequency	Cumulative frequency	Relative frequency	Cumulative relative frequency	Percent	Cumulative percent	Valid percent	Cumulative valid percent
0	388	388	0.91	0.91	90.65	90.65	90.65	90.65
1	40	428	0.09	1.00	9.35	100.00	9.35	100.00
Missing	0		0.00		0.00			
Total	428		1.00		100.00		100.00	

hyp - 1

Level	Frequency	Cumulative frequency	Relative frequency	Cumulative relative frequency	Percent	Cumulative percent	Valid percent	Cumulative valid percent
0	52	52	0.72	0.72	72.22	72.22	72.22	72.22
1	20	72	0.28	1.00	27.78	100.00	27.78	100.00
Missing	0		0.00		0.00			
Total	72		1.00		100.00		100.00	

[Normality test]

- 양적변수로 선택된 변수별로 정규성검정한 결과입니다.
- 표본수가 1,000개 미만인 경우는 Shapiro-Wilk test의 결과를, 1,000개 이상인 경우는 Kolmogorov-Smirnov test의 결과를 해석하는 편입니다.
- $p < 0.05$ 이면 데이터가 정규분포를 따르지 않는다고 판단합니다.

[Categorical Variables]

- [출력옵션] > [통계량] > [양적변수]에서 지정된 통계량이 변수별로 출력됩니다.

• 예제 - 결과창

Summary Table

Summary table (group: hyp)

	Total (N=500)	0 (N=428)	1 (N=72)	P-value
bweight	3188.5 (2862, 3551.25)	3252 (2917, 3571)	2980.5 (2414.5, 3220.5)	<0.001
gestwks	39.12 (37.94, 40.09)	39.35 (38.15, 40.17)	38.03 (36.97, 39.39)	<0.001
matage	34 (31, 37)	34 (32, 37)	33 (31, 37)	0.1555
lowbw				
- 0	440 (88%)	388 (90.65%)	52 (72.22%)	<0.001
- 1	60 (12%)	40 (9.35%)	20 (27.78%)	
preterm				
- 0	427 (87.14%)	375 (89.5%)	52 (73.24%)	<0.001
- 1	63 (12.86%)	44 (10.5%)	19 (26.76%)	
sex				
- 1	264 (52.8%)	221 (51.64%)	43 (59.72%)	0.2526
- 2	236 (47.2%)	207 (48.36%)	29 (40.28%)	

• This table was summarized as appropriate according to normality test and the presence of warning from chi-squared test.

Summary table in detail (group: hyp)

	Total (N=500) mean±SD (or n(%))	0 (N=428) mean±SD (or n(%))	1 (N=72) mean±SD (or n(%))	P- value <sup>1</sup>	Total (N=500) median(IQR) (or n(%))	0 (N=428) median(IQR) (or n(%))	1 (N=72) median(IQR) (or n(%))	P- value <sup>2</sup>	Normality test chi-squared or Fisher	Range (Total)	Range (0)	Range (1)
bweight	3136.88 ± 637.45	3198.9 ± 586.59	2768.21 ± 790.84	<0.001	3188.5 (2862, 3551.25)	3252 (2917, 3571)	2980.5 (2414.5, 3220.5)	<0.001	0, <0.001 / 1, 0.0033	628 to 4553	628 to 4553	693 to 4436
gestwks	38.72 ± 2.31	38.93 ± 2.09	37.51 ± 3.08	<0.001	39.12 (37.94, 40.09)	39.35 (38.15, 40.17)	38.03 (36.97, 39.39)	<0.001	0, <0.001 / 1, <0.001	24.69 to 43.16	24.69 to 43.16	27.33 to 42.2
matage	34.03 ± 3.9	34.13 ± 3.81	33.42 ± 4.38	0.1955	34 (31, 37)	34 (32, 37)	33 (31, 37)	0.1555	0, <0.001 / 1, 0.4195	23 to 43	23 to 43	24 to 43
lowbw									Chi-squared test			
- 0	440 (88%)	388 (90.65%)	52 (72.22%)	<0.001	440 (88%)	388 (90.65%)	52 (72.22%)	<0.001				
- 1	60 (12%)	40 (9.35%)	20 (27.78%)		60 (12%)	40 (9.35%)	20 (27.78%)					
preterm									Chi-squared test			
- 0	427 (87.14%)	375 (89.5%)	52 (73.24%)	<0.001	427 (87.14%)	375 (89.5%)	52 (73.24%)	<0.001				
- 1	63 (12.86%)	44 (10.5%)	19 (26.76%)		63 (12.86%)	44 (10.5%)	19 (26.76%)					
sex									Chi-squared test			
- 1	264 (52.8%)	221 (51.64%)	43 (59.72%)	0.2526	264 (52.8%)	221 (51.64%)	43 (59.72%)	0.2508				
- 2	236 (47.2%)	207 (48.36%)	29 (40.28%)		236 (47.2%)	207 (48.36%)	29 (40.28%)					

• Normality test was performed by Shapiro-Wilk test (3 ≤ sample size < 5000) or Kolmogorov-Smirnov test (sample size ≥ 5000).

• P-value<sup>1</sup>=Two sample t-test, P-value<sup>2</sup>=Wilcoxon rank sum test (Mann-Whitney U test) for continuous variable

• P-value<sup>1</sup>=Chi-squared test, P-value<sup>2</sup>=Fisher's exact test for categorical variable

Used R Packages

- Mean : 'mean' of R package 'base'
- Median : 'median' of R package 'stats'
- Standard deviation : 'sd' of R package 'stats'
- Variance : 'var' of R package 'stats'
- Skewness : 'skewness' of R package 'moments'
- Kurtosis : 'kurtosis' of R package 'moments'
- All results other than those mentioned above were written with basic functions of R.

[Summary Table]

[Summary table (group: hyp)]

- [집단변수]로 지정된 hyp에 따른 변수들의 요약표가 출력됩니다.
- 양적변수의 경우 정규성 검정 결과에 따라 정규분포를 따르는 경우 평균과 표준편차로, 정규분포를 따르지 않는 경우 중앙값과 사분위수 범위로 요약됩니다. 정규분포를 따르는 경우 2개 수준의 집단 비교 시 Student's t-test, 3개 이상 수준의 집단 비교 시 one-way ANOVA 의 p-value가 출력되며, 정규분포를 따르지 않는 경우 2개 수준의 집단 비교 시 Wilcoxon's ranksum test, 3개 이상 수준의 집단 비교 시 Kruskal-Wallis test의 p-value가 출력됩니다.
- 질적변수의 경우 빈도와 열백분율(%)로 요약됩니다. 기본적으로 카이제곱검정 결과 p-value가 출력되며, 카이제곱검정의 가정에 위배되는 경우 Fisher's exact test의 p-value가 출력됩니다. Fisher's exact test가 수행되기 어려운 경우, 가정에 위배되더라도 카이제곱검정의 p-value가 출력됩니다.

[Summary table in detail (group: hyp)]

- [집단변수]로 지정된 hyp에 따른 변수들의 요약표가 출력됩니다.
- 정규성 검정 결과와 무관하게, (1) 모수적 분포를 가정한 결과, (2) 비모수적 방식으로 계산된 결과 순으로 결과가 출력됩니다.
- [Normality test / Chi-squared of Fisher] 에서 해당 변수가 양적변수인 경우 집단 수준별 정규성 검정 결과 p-value가 출력되고, 해당변수가 질적변수인 경우는 카이제곱검정이 수행되었는지, Fisher's exact test가 수행되었는지를 구분합니다.
- [Range]에서는 양적변수의 경우 최소값과 최대값이 출력됩니다.

[Used R Packages]

- 본 분석에 사용된 R 패키지들이 나열되어 있습니다.