

EXAMINE VARIABLES=왕수명 /PLOT BOXPLOT STEMLEAF /COMPARE GROUP /STATISTICS  
 DESCRIPTIVES /CINTERVAL 95 /MISSING LISTWISE /NOTOTAL.

## Explore

### Notes

	Output Created	17:5월:201110시 28분 03초
	Comments	
Input	Active Dataset	DataSet0
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data	27
	File	
Missing Value Handling	Definition of Missing	User-defined missing values for dependent variables are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any dependent variable or factor used.
	Syntax	EXAMINE VARIABLES=왕수명 /PLOT BOXPLOT STEMLEAF /COMPARE GROUP /STATISTICS DESCRIPTIVES /CINTERVAL 95 /MISSING LISTWISE /NOTOTAL.
Resources	Processor Time	0:00:01.357
	Elapsed Time	0:00:01.856

[DataSet0]

### Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
왕수명	27	100.0%	0	.0%	27	100.0%

**Descriptives**

		Statistic	Std. Error
왕수명	Mean	46.0370	3.11255
95% Confidence Interval for Mean	Lower Bound	39.6391	
	Upper Bound	52.4350	
	5% Trimmed Mean	45.7510	
	Median	45.0000	
	Variance	261.575	
	Std. Deviation	16.17330	
	Minimum	16.00	
	Maximum	82.00	
	Range	66.00	
	Interquartile Range	23.00	
	Skewness	.274	.448
	Kurtosis	-.420	.872

평균과 표준편차를 봐서 평균사망나이는 46세이며 표준편차는 16년으로 다소 편차가 크다고 볼 수 있다.

**왕수명**

왕수명에 대한 줄기-잎 그림의 경우 30세와 50세에 분포가 집중되어 있는 쌍봉형 형태임을 알 수 있다.

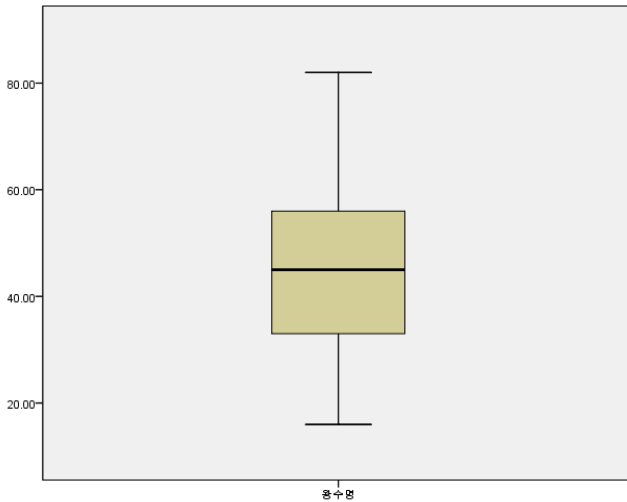
왕수명 Stem-and-Leaf Plot

```

Frequency   Stem & Leaf
1.00        1 . 6
2.00        2 . 28
8.00        3 . 00233678
4.00        4 . 0458
7.00        5 . 1234669
3.00        6 . 267
1.00        7 . 3
    
```

1.00        8 . 2  
Stem width:    10.00  
Each leaf:     1 case(s)

boxplot결과로 볼때는 중위수값과 평균 차이가 없으며 1사분위수와 3사분위수 각각 중위수에서의 거리를 볼 때 균일한 것으로 봐서 대략적으로 정규분포의 형태를 띄고 있음을 알 수 있다.



```
EXAMINE VARIABLES=환자나이 /PLOT BOXPLOT STEMLEAF /COMPARE GROUP  
/PERCENTILES(5,10,25,50,75,90,95) HAVERAGE /STATISTICS DESCRIPTIVES  
/CINTERVAL 95 /MISSING LISTWISE /NOTOTAL.
```

## Explore

### Notes

Input	Output Created	17:5월:201111시 27분 06초
	Comments	
	Data	D:\knou\통계패키지\5-6.sav
	Active Dataset	DataSet0
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data	100
	File	

Missing Value Handling	Definition of Missing	User-defined missing values for dependent variables are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any dependent variable or factor used.
	Syntax	EXAMINE VARIABLES=환자나이 /PLOT BOXPLOT STEMLEAF /COMPARE GROUP /PERCENTILES(5,10,25,50,75,90,95) HAVERAGE /STATISTICS DESCRIPTIVES /CINTERVAL 95 /MISSING LISTWISE /NOTOTAL.
Resources	Processor Time	0:00:00.468
	Elapsed Time	0:00:00.491

[DataSet0] D:\knou\통계패키지\5-6.sav

**Case Processing Summary**

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
환자나이	100	100.0%	0	.0%	100	100.0%

회사의 평균 나이는 34.6세 표준편차는 9.7세이다.

**Descriptives**

		Statistic	Std. Error
환자나이	Mean	34.63	.978
95% Confidence Interval for Mean	Lower Bound	32.69	
	Upper Bound	36.57	
	5% Trimmed Mean	34.14	
	Median	33.00	

Variance	95.710	
Std. Deviation	9.783	
Minimum	17	
Maximum	63	
Range	46	
Interquartile Range	13	
Skewness	.742	.241
Kurtosis	.287	.478

중위수가 평균보다 약간 작은 33세인걸로 보서는 33세 이하 나이의 연령 분포가 작고 33세 이상의 연령 분포가 큰 것으로 예상할 수 있다.

**Percentiles**

	Percentiles				
	5	10	25	50	75
Weighted Average(Definition 환자나이 1)	21.00	23.10	27.25	33.00	40.00
Tukey's Hinges 환자나이			27.50	33.00	40.00

**Percentiles**

	Percentiles	
	90	95
Weighted Average(Definition 환자나이 1)	49.90	54.90

## 환자나이

줄기 잎 그림으로 볼 때 분포는 30세 초반을 중심으로 단봉형 분포를 보이고 있고 나이가 많은 쪽으로 분포가 길게 늘어지는 모습을 볼 수 있다.

환자나이 Stem-and-Leaf Plot

```

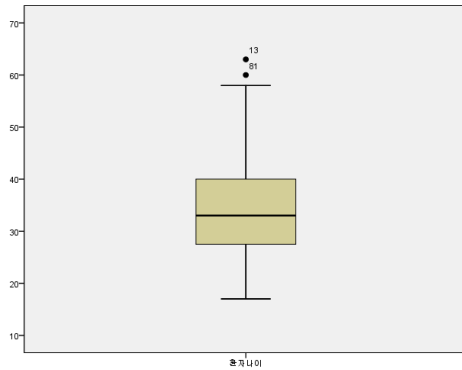
Frequency    Stem & Leaf
 2.00        1 . 78
10.00        2 . 0111223344
19.00        2 . 5555555666667888999
27.00        3 . 00000111122223333333344444444
16.00        3 . 55556666667888999
10.00        4 . 0001123334
    
```

<http://freesearch.pe.kr>

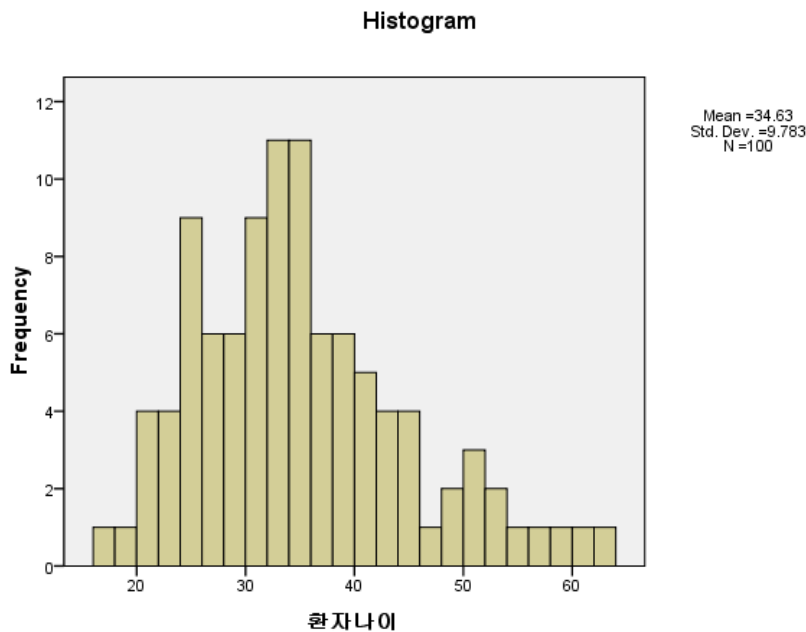
```
6.00      4 . 555689
5.00      5 . 01123
3.00      5 . 578
2.00 Extremes (>=60)
```

```
Stem width: 10
Each leaf: 1 case(s)
```

boxplot결과를 볼 때 outlier로 13번째 81번째 데이터를 볼 수 있으며 이들 데이터 값은 63세와 60세이다.



히스토그램은 줄기-잎 그림과 비슷한 분포를 보여주나, 2세 단위로 끊어진 x값 덕분에 좀더 자세한 분포를 볼 수 있다.



```
WEIGHT BY 사원수. CROSSTABS /TABLES=교육수준 BY 흡연실태 /FORMAT=AVALUE TABLES
/STATISTICS=CHISQ /CELLS=COUNT /COUNT ROUND CELL.
```

## Crosstabs

### Notes

	Output Created	17:5월:201114시 12분 00초
	Comments	
Input	Active Dataset	DataSet0
	Filter	<none>
	Weight	사원수
	Split File	<none>
	N of Rows in Working Data File	9
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each table are based on all the cases with valid data in the specified range(s) for all variables in each table.
	Syntax	CROSSTABS /TABLES=교육수준 BY 흡연실태 /FORMAT=AVALUE TABLES /STATISTICS=CHISQ /CELLS=COUNT /COUNT ROUND CELL.
Resources	Processor Time	0:00:00.015
	Elapsed Time	0:00:00.017
	Dimensions Requested	2
	Cells Available	174762

[DataSet0]

### Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent

**Case Processing Summary**

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
교육수준 * 흡연상태	356	100.0%	0	.0%	356	100.0%

**교육수준 \* 흡연상태 Crosstabulation**

Count

		흡연상태			Total
		과흡연	비흡연	흡연	
교육수준	고졸	22	9	21	52
	대졸	51	68	92	211
	중졸	43	22	28	93
	Total	116	99	141	356

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	18.510 <sup>a</sup>	4	.001
Likelihood Ratio	18.663	4	.001
N of Valid Cases	356		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 14.46.

귀무가설 : 흡연상태와 교육수준 변수는 상호 독립이다.

대립가설 : 흡연상태와 교육수준 변수는 상호 독립이 아니다.

테스트 결과 p-value는 0.001로 유의한 값이 나왔으며 유의수준 0.05로 검정할 때 대립가설을 지지하게 된다. 따라서 흡연상태와 교육수준 변수는 상호 독립이 아니다.

T-TEST GROUPS=그룹 (1 2) /MISSING=ANALYSIS /VARIABLES=상표  
/CRITERIA=CI (.95).

**T-Test**



**Notes**

	Output Created	17:5월:201117시 59분 19초
	Comments	
Input	Active Dataset	DataSet0
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	21
Missing Value Handling	Definition of Missing	User defined missing values are treated as missing.
	Cases Used	Statistics for each analysis are based on the cases with no missing or out-of-range data for any variable in the analysis.
	Syntax	T-TEST GROUPS=그룹(1 2) /MISSING=ANALYSIS /VARIABLES=상표 /CRITERIA=CI(.95).
Resources	Processor Time	0:00:00.016
	Elapsed Time	0:00:00.012

[DataSet0]

**Group Statistics**

그룹	N	Mean	Std. Deviation	Std. Error Mean
상표 1	11	80.518	2.0380	.6145
2	10	79.150	1.6595	.5248

**Independent Samples Test**

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
상표	Equal variances assumed	.040	.844	1.676	19
	Equal variances not assumed			1.693	18.797

**Independent Samples Test**

		t-test for Equality of Means		
		Sig. (2-tailed)	Mean Difference	Std. Error Difference
상표	Equal variances assumed	.110	1.3682	.8163
	Equal variances not assumed	.107	1.3682	.8081

**Independent Samples Test**

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
상표	Equal variances assumed	-.3404	3.0768
	Equal variances not assumed	-.3244	3.0608

먼저 등분산 가정이 성립 되는지 등분산 테스트 결과를 확인해 보면 상표1,2간 등분산 가정이 성립됨을 유의확률 0.844로 확인 할 수 있다. 따라서 이 데이터의 t-test 검정 결과는 등분산이 가정된 경우(Equal variances assumed)항목으로 확인 가능하다. 표에서 유의확률 0.110으로 유의수준 0.05보다 큰 값을 가지므로 두 집단의 평균의 차이가 없다는 귀무가설을 기각할 수 없다.

T-TEST PAIRS=사용전 WITH 사용후 (PAIRED) /CRITERIA=CI (.9500)  
/MISSING=ANALYSIS.

**T-Test**

**Notes**

	Output Created	17:5월:201118시 25분 56초
	Comments	
Input	Active Dataset	DataSet0
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	15
Missing Value Handling	Definition of Missing	User defined missing values are treated as missing.
	Cases Used	Statistics for each analysis are based on the cases with no missing or out-of-range data for any variable in the analysis.
	Syntax	T-TEST PAIRS=사용전 WITH 사용후 (PAIRED) /CRITERIA=CI(.9500) /MISSING=ANALYSIS.
Resources	Processor Time	0:00:00.015
	Elapsed Time	0:00:00.025

[DataSet0]

**Paired Samples Statistics**

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 사용전	75.87	15	6.865	1.772
사용후	67.07	15	6.670	1.722

**Paired Samples Correlations**

	N	Correlation	Sig.
Pair 1 사용전 & 사용후	15	-.315	.253

**Paired Samples Test**

		Paired Differences		
		Mean	Std. Deviation	Std. Error Mean
Pair 1	사용전 - 사용후	8.800	10.975	2.834

**Paired Samples Test**

		Paired Differences				
		95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
		Lower	Upper			
Pair 1	사용전 - 사용후	2.722	14.878	3.105	14	.008

```
SAVE TRANSLATE OUTFILE='D:\knou\pkg\6-4.csv' /TYPE=CSV /MAP /REPLACE
/FIELDNAMES /CELLS=VALUES. Data written to D:\knou\pkg\6-4.csv. 2 variables
and 15 cases written. Variable: 사용전 Type: Number Width: 8 Dec:
0 Variable: 사용후 Type: Number Width: 8 Dec: 0
```

H0: 피임약 사용이 혈압에 영향을 주지 않는다.

H1: 피임약 사용이 혈압을 저하 시킨다.

유의확률이 양측검정시 0.008이 나왔는데, 단측검정은 0.004로 유의수준 0.05보다 작은 값이 나온다. 따라서 피임약 사용은 혈압을 저하 시킨다는 대립가설을 지지하게 된다.

```
T-TEST PAIRS=사용전 WITH 사용후 (PAIRED) /CRITERIA=CI(.9500)
/MISSING=ANALYSIS.
```

**T-Test**

**Notes**

	Output Created	17:5월:201120시 45분 50초
	Comments	
Input	Active Dataset	DataSet1

	Filter	<none>	
	Weight	<none>	
	Split File	<none>	
	N of Rows in Working Data		10
	File		
Missing Value Handling	Definition of Missing	User defined missing values are treated as missing.	
	Cases Used	Statistics for each analysis are based on the cases with no missing or out-of-range data for any variable in the analysis.	
	Syntax	T-TEST PAIRS=사용전 WITH 사용후 (PAIRED) /CRITERIA=CI(.9500) /MISSING=ANALYSIS.	
Resources	Processor Time		0:00:00.015
	Elapsed Time		0:00:00.010

[DataSet1]

**Paired Samples Statistics**

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 사용전	13.660	10	.5816	.1839
사용후	13.490	10	.5301	.1676

**Paired Samples Correlations**

	N	Correlation	Sig.
Pair 1 사용전 & 사용후	10	.921	.000

**Paired Samples Test**

Paired Differences	

	Mean	Std. Deviation	Std. Error Mean
Pair 1 사용전 - 사용후	.1700	.2263	.0716

**Paired Samples Test**

	Paired Differences				
	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
	Lower	Upper			
Pair 1 사용전 - 사용후	.0081	.3319	2.375	9	.042

H0 : 감량 전후 속도 차이가 없다.

H1 : 감량 전후 속도 차이가 있다.

양측검정으로 테스트 결과의 유의확률 값을 그대로 사용하면 된다. 표에서 유의확률 0.042로 유의수준 0.05보다 작은 값을 가지므로 귀무가설을 기각하고 대립가설을 선택하여 감량 전후 속도 차이는 있다고 결론지을 수 있다.

```
* Visual Binning. *기계부품길이. RECODE 기계부품길이 (MISSING=COPY) (LO THRU 15.75=1) (LO THRU 15.8850000000000=2) (LO THRU 15.94=3) (LO THRU 15.96=4) (LO THRU 16.03=5) (LO THRU 16.09=6) (LO THRU 16.18=7) (LO THRU HI=8) (ELSE=SYSMIS) INTO bin. VARIABLE LABELS bin '기계부품길이 (Binned)'. FORMAT bin (F5.0). VALUE LABELS bin 1 '= 15.75' 2 '15.76 - 15.89' 3 '15.90 - 15.94' 4 '15.95 - 15.96' 5 '15.97 - 16.03' 6 '16.04 - 16.09' 7 '16.10 - 16.18' 8 '16.19+'. MISSING VALUES bin (.). VARIABLE LEVEL bin (ORDINAL). EXECUTE. IGRAPH /VIEWNAME='Bar Chart' /X1=VAR(bin) TYPE=CATEGORICAL /Y=$count /COORDINATE=VERTICAL /YLENGTH=5.2 /X1LENGTH=6.5 /CHARTLOOK='NONE' /CATORDER VAR(bin) (ASCENDING VALUES OMITEMPTY) /BAR KEY=ON SHAPE=RECTANGLE BASELINE=AUTO.
```

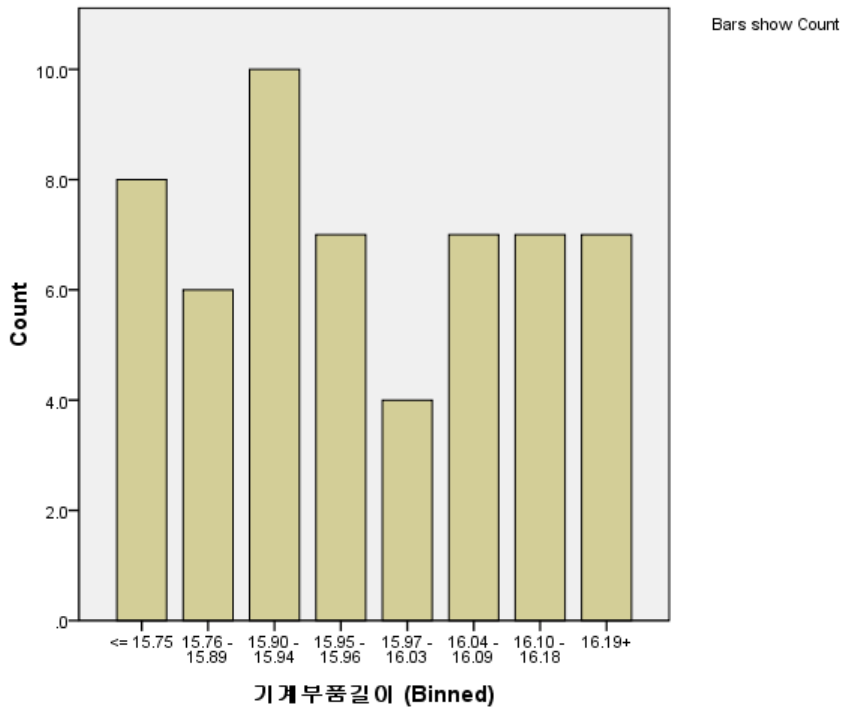
**Interactive Graph**

**Notes**

Output Created	18:5월:201100시 29분 06초
Comments	
Input Data	D:\knou\pkg\7-4.sav
Active Dataset	DataSet1
Filter	<none>
Weight	<none>

Split File	<none>
N of Rows in Working Data	56
File	
Syntax	<pre>IGRAPH   /VIEWNAME='Bar Chart'   /X1=VAR(bin) TYPE=CATEGORICAL   /Y=\$count   /COORDINATE=VERTICAL   /YLENGTH=5.2   /X1LENGTH=6.5   /CHARTLOOK='NONE'   /CATORDER VAR(bin) (ASCENDING VALUES OMITEMPTY)   /BAR KEY=ON SHAPE=RECTANGLE BASELINE=AUTO.</pre>
Resources	
Processor Time	0:00:00.452
Elapsed Time	0:00:00.492

[DataSet1] D:\knou\pkg\7-4.sav



```
IGRAPH /VIEWNAME='Bar Chart' /X1=VAR(연도) TYPE=CATEGORICAL
/Y=VAR(투표율) TYPE=SCALE /COORDINATE=VERTICAL /YLENGTH=5.2
/X1LENGTH=6.5 /CHARTLOOK='NONE' /CATORDER VAR(연도) (ASCENDING VALUES
OMITEMPTY) /BAR(MEAN) KEY=ON SHAPE=RECTANGLE BASELINE=AUTO.
```

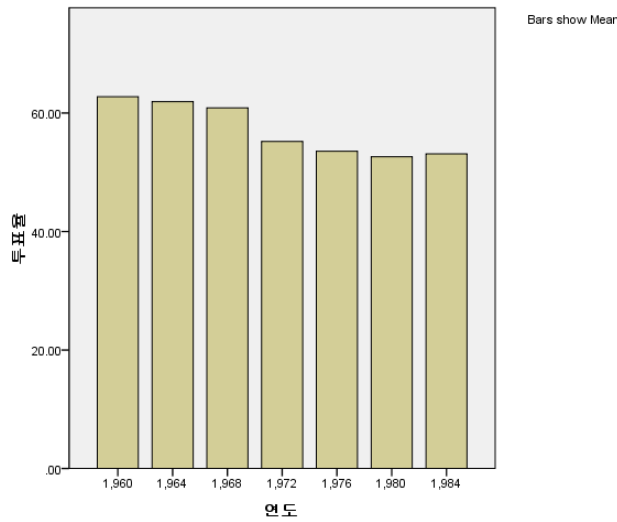
## Interactive Graph

### Notes

	Output Created	18:5월:201115시 51분 51초
	Comments	
Input	Data	D:\knou\pkg\7-5.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data	7
	File	
	Syntax	IGRAPH /VIEWNAME='Bar Chart' /X1=VAR(연도) TYPE=CATEGORICAL /Y=VAR(투표율) TYPE=SCALE /COORDINATE=VERTICAL /YLENGTH=5.2 /X1LENGTH=6.5 /CHARTLOOK='NONE' /CATORDER VAR(연도) (ASCENDING VALUES OMITEMPTY) /BAR(MEAN) KEY=ON SHAPE=RECTANGLE BASELINE=AUTO.
Resources	Processor Time	0:00:00.484
	Elapsed Time	0:00:00.546



[DataSet2] D:\knou\pkg\7-5.sav



투표율 변수를 유권자수와 투표자수 데이터를 이용해서 만들고 이를 막대 그래프로 표현한 결과이다. 시간이 가면 갈수록 투표율은 점점 줄어드는 경향을 보이는 것을 그래프로 알 수 있으나 1984년이 1980년도의 투표율보다 다소 증가한 모습을 보이고 있다.

```
GET FILE='D:\knou\pkg\7-5.sav'. CORRELATIONS /VARIABLES=홍런수 평균타율  
/PRINT=TWOTAIL NOSIG /MISSING=PAIRWISE.
```

## Correlations

### Notes

	Output Created	22:5월:201109시 03분 50초
	Comments	
Input	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data	14
	File	
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.

	Cases Used	Statistics for each pair of variables are based on all the cases with valid data for that pair.
	Syntax	CORRELATIONS /VARIABLES=홈런수 평균타율 /PRINT=TWOTAIL NOSIG /MISSING=PAIRWISE.
Resources	Processor Time	0:00:00.000
	Elapsed Time	0:00:00.061

[DataSet2]

**Correlations**

		홈런수	평균타율
홈런수	Pearson Correlation	1	-.075
	Sig. (2-tailed)		.800
	N	14	14
평균타율	Pearson Correlation	-.075	1
	Sig. (2-tailed)	.800	
	N	14	14

상관행렬을 보면 상관계수는 -0.075로 0에 가까운 상관관계를 보여 별다른 상관관계가 없다는 것을 알 수 있으며 상관계수가 0이라는 귀무가설로 검정한 p-value가 0.800으로 0.05보다 큰 값을 보여 귀무가설을 지지하게 된다. 따라서 평균 타율과 홈런수는 상관관계가 없다는 결론을 내릴 수 있다.

\* Curve Estimation. TSET NEWVAR=NONE. CURVEFIT /VARIABLES=강도 WITH 목재길이  
/CONSTANT /MODEL=LINEAR /PRINT ANOVA /PLOT FIT.

**Curve Fit**

**Notes**

Output Created	22:5월:201109시 18분 17초
Comments	

Input	Data	D:\knou\pkg\8-7.sav
	Active Dataset	DataSet0
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	10
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Cases with a missing value in any variable are not used in the analysis.
	Syntax	CURVEFIT /VARIABLES=강도 WITH 목재길이 /CONSTANT /MODEL=LINEAR /PRINT ANOVA /PLOT FIT.
Resources	Processor Time	0:00:00.483
	Elapsed Time	0:00:00.503
Use	From	First observation
	To	Last observation
Predict	From	First Observation following the use period
	To	Last observation
Time Series Settings (TSET)	Amount of Output	PRINT = DEFAULT
	Saving New Variables	NEWVAR = NONE
	Maximum Number of Lags in Autocorrelation or Partial Autocorrelation Plots	MXAUTO = 16
	Maximum Number of Lags Per Cross-Correlation Plots	MXCROSS = 7
	Maximum Number of New Variables Generated Per Procedure	MXNEWVAR = 60
	Maximum Number of New Cases Per Procedure	MXPREDICT = 1000
	Treatment of User-Missing Values	MISSING = EXCLUDE

Confidence Interval	CIN = 95
Percentage Value	
Tolerance for Entering	TOLER = .0001
Variables in Regression	
Equations	
Maximum Iterative Parameter Change	CNVERGE = .001
Method of Calculating Std. Errors for Autocorrelations	ACFSE = IND
Length of Seasonal Period	Unspecified
Variable Whose Values Label	Unspecified
Observations in Plots	
Equations Include	CONSTANT

[DataSet0] D:\knou\pkg\8-7.sav

**Model Description**

Model Name	MOD_2
Dependent Variable	1 강도
Equation	1 Linear
Independent Variable	목재길이
Constant	Included
Variable Whose Values Label	Unspecified
Observations in Plots	

**Case Processing Summary**

	N
Total Cases	10
Excluded Cases <sup>a</sup>	0
Forecasted Cases	0
Newly Created Cases	0

a. Cases with a missing value in any variable are excluded from the analysis.

**Variable Processing Summary**

		Variables	
		Dependent	Independent
		강도	목재길이
	Number of Positive Values	10	10
	Number of Zeros	0	0
	Number of Negative Values	0	0
Number of Missing Values	User-Missing	0	0
	System-Missing	0	0

**강도**

**Linear**

**Model Summary**

R	R Square	Adjusted R Square	Std. Error of the Estimate
.882	.778	.751	35.553

The independent variable is 목재길이.

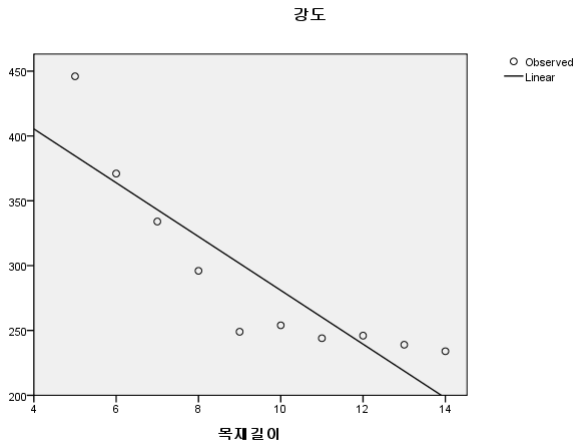
**ANOVA**

	Sum of Squares	df	Mean Square	F	Sig.
Regression	35505.845	1	35505.845	28.089	.001
Residual	10112.255	8	1264.032		
Total	45618.100	9			

The independent variable is 목재길이.

**Coefficients**

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
목재길이	-20.745	3.914	-.882	-5.300	.001
(Constant)	488.382	38.848		12.572	.000



R Square가 0.778로 회귀식이 설명력이 크다고 이야기 할 수 있으며 목재길기와 강도의 관계의 연관성이 크다는 것을 알 수 있다. 게다가 ANOVA로 확인한 목재길이 변수의 유의성에 대한 p-value는 0.001로 0.05보다 작아 유의하다고 볼 수 있다.

회귀식은 아래와 같다.

$$\text{강도} = \text{목재길이} * (-20.745) + 488.382$$

강도가 가장 큰 목재를 구하려면 목재길이가 0이 되어야 하나 목재 길이가 0인 목재는 없으므로 1이상의 값중에서 적절한 강도에 대한 목재길이를 정하면 될 것이다.