

이번에는 Histogram 에 대해 알아보겠습니다. 통계적 분석을 하기 전에, 반드시 Graph 분석을 실시하여 분석 대상 Data 의 전체적인 모습, 연관성, 경향 등을 파악하여야 하는 데, Histogram 은 그 중에서도 가장 기본적인 Graph 라 할 수 있습니다

Histogram 실습 1

1. 평균 50, 표준편차 5, Spec (48 ~ 58) 인 A 라는 Data 에 대해 난수(Random Data)를 발생시킨 후, Histogram 을 그려보겠습니다(정규 분포 가정)

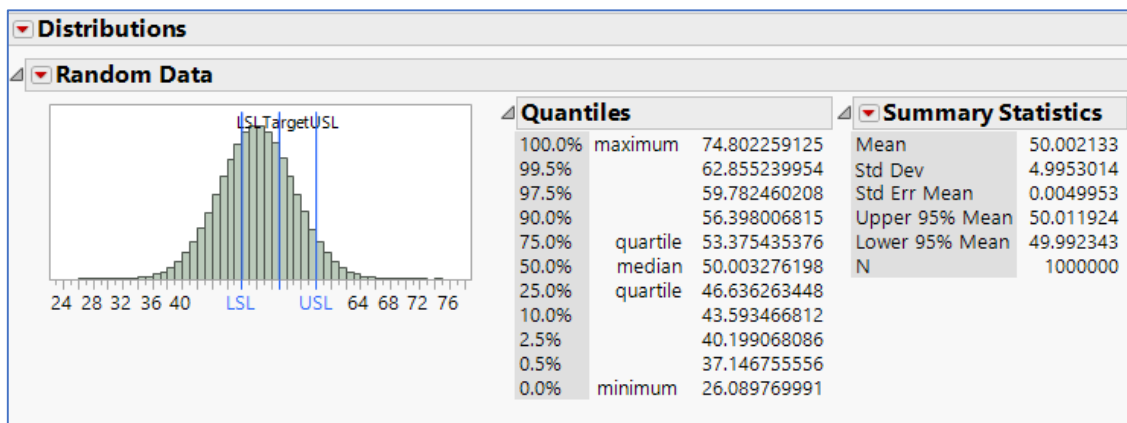
1) 난수 발생 : cols / new column

(다른 방법 : data file 의 해당 column 에서 오른쪽 마우스 click 후 'column info' 선택)

'Column Properties' 에서 Spec Limit 선택 후 Spec 입력

2. Analyze / Distribution

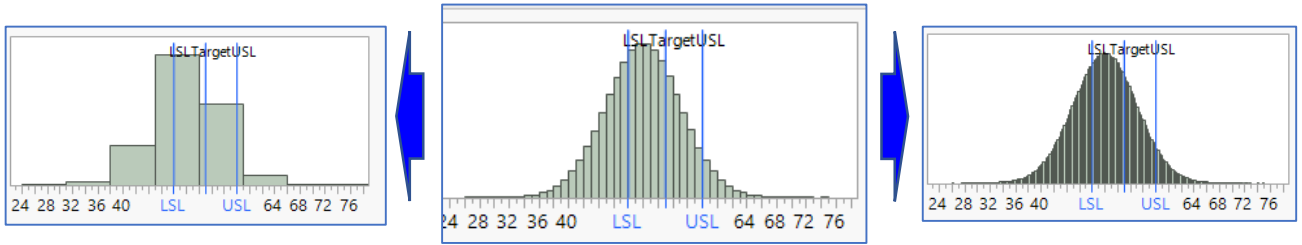
1) 'Distribution' 의 왼쪽 역삼각형 click 하여 'stack' 을 선택하면 아래와 같이 가로로 배열됨



-Uniform Scaling : 여러 개의 Histogram 을 그릴 때 척도(Scale) 를 동일하게 할 경우

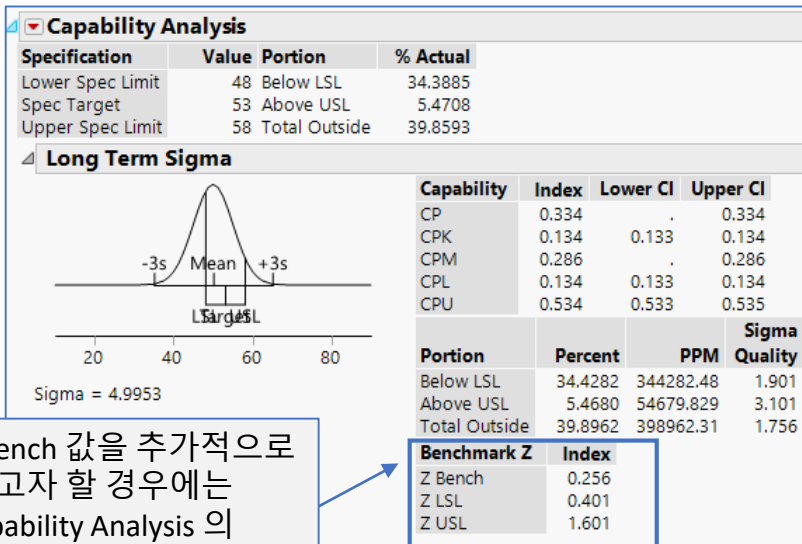
-Arrange in Row : 여러 개의 Histogram 을 그릴 때 한 행(가로)에 display 할 Histogram 의 수 (예를 들어, Histogram 10 개가 있을 경우 'Arrange in Row' 에 '5' 를 입력하면 Histogram 이 가로 5개, 세로 2개 display 됨)

- 2) 화면 상단의 Grabber (장갑처럼 생긴) Icon 선택 후 마우스를 Histogram 위에 놓은 후, 위아래로 움직이면 Histogram 이 모양이 아래와 같이 변함



- 3) 다른 통계량을 Display 시키고자 할 경우에는 'Summary Statistics' 에서 'Customize Summary Statistics' 를 click 하여 보고자 하는 통계량을 선택하면 됨

- 4) Spec 을 미리 입력하면 Histogram 이 Display 될 때, 공정 능력 분석이 함께 됨



Z bench 값을 추가적으로 보고자 할 경우에는 Capability Analysis 의 왼쪽 역삼각형을 click 한 후 'Z bench' 선택

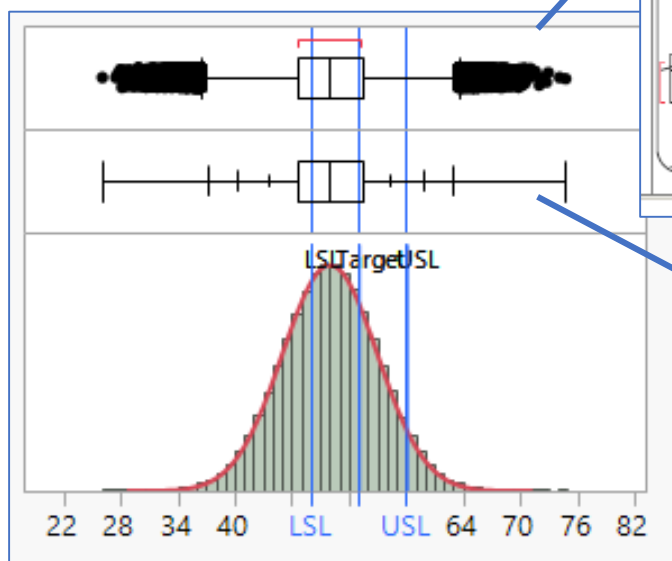
The figure shows the 'Summary Statistics Options' dialog box. It contains a list of statistical options with checkboxes. The 'Mean' option is checked. Other options include Std Dev, Std Err Mean, Upper Mean Confidence Limits, Lower Mean Confidence Limits, N, Sum Weight, Sum, Variance, Skewness, Kurtosis, CV, N Missing, N Zero, N Unique, Uncorrected SS, Corrected SS, Autocorrelation, Minimum, Maximum, Median, Mode, Trimmed Mean, Geometric Mean, Range, Interquartile Range, Median Absolute Deviation, Robust Mean, and Robust Std Dev. There are buttons for 'Select All Summary Statistics' and 'Deselect All Summary Statistics'. At the bottom, there are input fields for 'Enter (1-alpha) for mean confidence interval' (set to 0.95) and 'Enter trimmed mean percent' (set to 5).

- 5) 그 외 많이 활용되는 기능

- 2-1) 의 'Random Data' 의 모든 Option Menu 를 한꺼번에 보고자 할 경우에는 'Alt' key 를 누른 상태에서 왼쪽 역삼각형을 click 하면 아래의 예시처럼 하위 모든 Option 이 Display 됨

The figure shows the 'Display Options' dialog box. It contains a list of display options with checkboxes. The 'Quantiles' option is checked. Other options include Set Quantile Increment, Custom Quantiles, Summary Statistics, Customize Summary Statistics, Horizontal Layout, Axes on Left, Histogram Options, Shadowgram, Vertical, Std Error Bars, Set Bin Width, Count Axis, Prob Axis, Density Axis, Show Percents, Show Counts, Normal Quantile Plot, Outlier Box Plot, Quantile Box Plot, Stem and Leaf, CDF Plot, Test Mean, Test Std Dev, Test Equivalence, Confidence Interval, Prediction Interval, Tolerance Interval, Capability Analysis, Continuous Fit, GLog, All, LogNormal, Weibull, Weibull with Threshold, Extreme Value, Exponential, Gamma, Beta, Normal Mixtures, Normal 2 Mixture, Normal 3 Mixture, Other, Smooth Curve, Johnson Su, Johnson Sb, and Johnson Sl. There are buttons for 'Save' and 'Level Numbers'.

6) Outlier Box plot & Quantile Box Plot



whisker
3rd quartile
shortest half
median sample value
confidence diamond
1st quartile
whisker

Whisker 계산

$1st\ quartile - 1.5 * (IQR)$

$3rd\ quartile + 1.5 * (IQR)$

$IQR(interquartile\ range)$

$= 3Q - 1Q$

Quantiles		
100.0%	maximum	74.802259125
99.5%		62.855239954
97.5%		59.782460208
90.0%		56.398006815
75.0%	quartile	53.375435376
50.0%	median	50.003276198
25.0%	quartile	46.636263448
10.0%		43.593466812
2.5%		40.199068086
0.5%		37.146755556
0.0%	minimum	26.089769991

7) Confidence Interval, Tolerance Interval, Prediction Interval

Confidence Intervals				
Parameter	Estimate	Lower CI	Upper CI	1-Alpha
Mean	50.00213	49.99234	50.01192	0.950
Std Dev	4.995301	4.988388	.	0.950

Tolerance Intervals			
Proportion	Lower TI	Upper TI	1-Alpha
0.900	41.78559	58.21867	0.950

Prediction Interval				
Parameter	Future N	Lower PI	Upper PI	1-Alpha
Individual	1	40.21151	59.79276	0.950
Mean	1	40.21151	59.79276	0.950
Std Dev	1	.	.	0.950

Confidence Intervals

: 모집단 평균의 신뢰 구간

Tolerance interval

: 해당 비율 만큼의

Data 가 존재할 최소 구간

Prediction intervals

: 모집단의

특정 data 가 존재할 구간

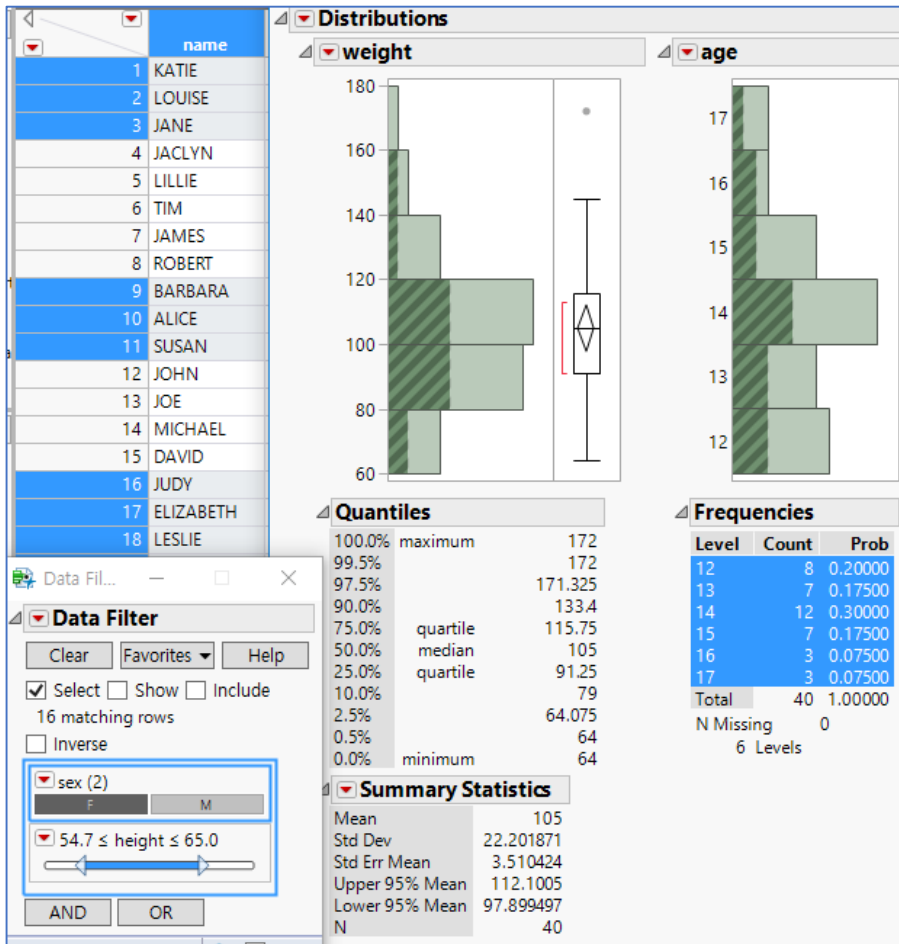
Histogram 실습 2

예제 파일 : big class.jmp

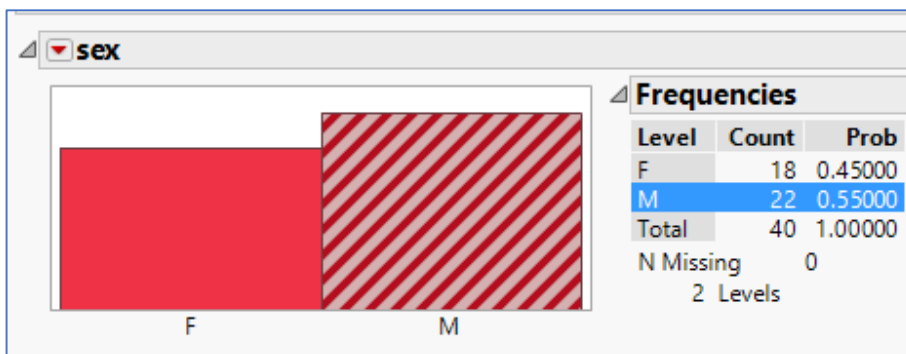
1. Analyze / distribution 에서 weight, age 선택 후 OK Click
2. Rows / data filter 에서 sex, height 를 선택 후 Add click

3. 결과

- 1) Data Filter 의 설정(일종의 조건 설정) 결과에 따라 Histogram 간 및 Histogram 과 data 가 연동되어 display 됨을 알 수 있다 : Interactive !!!
- 2) Age 와 같은 Ordinal Data 는 Level 별 빈도 수를 보여줌



3) Nominal Data 또한 Level 별 빈도 수를 display 함



* 본 Guide 는 매월 네 번째 수요일에 발행됩니다

** Monthly User Guide 지난 호는 다음 Site 를 참조하세요.

https://www.jmp.com/ko_kr/newsletters.html

*** 본 Guide 의 내용과 관련한 문의는 Ikju.Shin@jmp.com 으로 연락 바랍니다