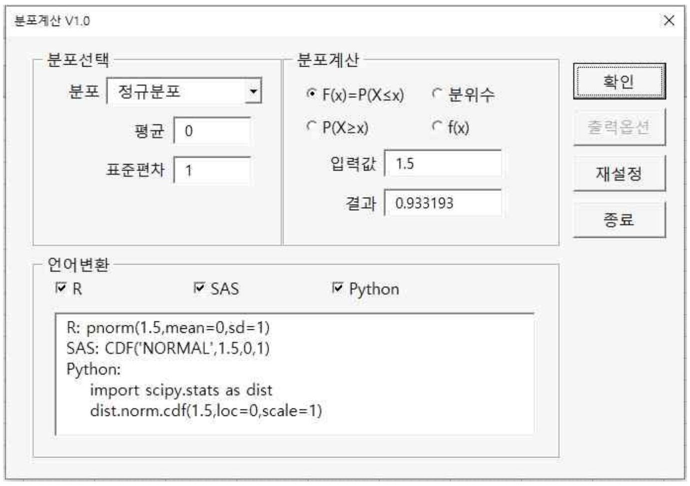
◉ 분포계산

○ 분포 및 추론 > 분포계산

- 다양한 표준분포의 CDF($F(x)=P(X \le x)$), 분위수($P(X \le x)=\alpha$를 만족하는 x의 최소값), $P(X \ge x)$의 값, PDF값을 계산해 줌
- 분포선택에서 모수를 지정하고 분포계산에서 계산하고자 하는 함수를 선택한 후 입력값에 값을 입력한 후 확인 선택
- 이산분포(이항분포, 초기하분포, 음이항분포) 모수 중 횟수나 크기에 해당하는 모수가 자연수가 아닌 경우에는 경고메시지가 뜨고 실행되지 않음. 분위수의 경우 입력값은 0에서 1사이의 값이어야 하며 입력값이 0이하 1이상인 경우 경고메시지가 뜨고 실행되지 않음. 모수공간을 벗어난 모수가 지정된 경우에도 경고메시지가 뜨고 실행되지 않음
- 동일한 결과를 R, SAS, Python으로 구하고 싶은 경우 '언어변환'에서 해당 언어를 선택하면 됨

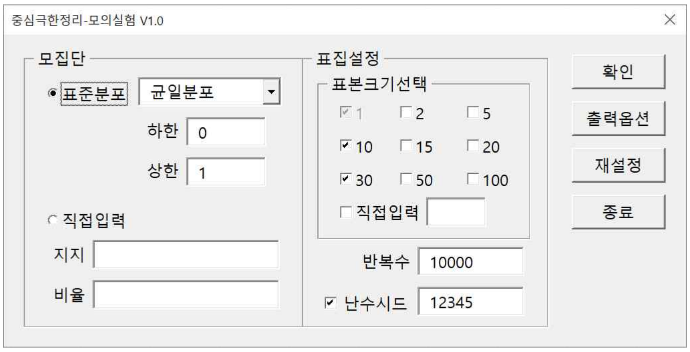


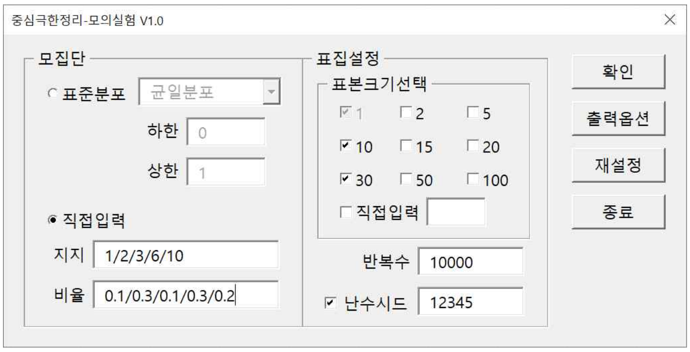
[그림 1] 분포계산 분석폼

◉ 중심극한정리

○ 분포 및 추론 > 중심극한정리

- ○ 모의실험을 통해 중심극한정리가 성립하는 것을 시각적으로 보여 줌
- ○ 모집단 프레임에서는 모집단에 대한 가정을 설정하는 것으로 표준분포와 직접입력하는 방식을 제공함. 직접입력의 경우 지지(support)와 해당 지지의 비율을 "/" 또는 ","로 구분하여 입력하며 비율의 합은 1일 필요가 없음.
- ○ 표집설정 프레임에서는 표본크기를 선택하거나 입력할 수 있으며 해당 표본크기의 표본평균을 반복수만큼 계산하고 결과를 히스토그램으로 표시함. 이론적인 정규분포를 같이 출력하여 비교함. 표본크기가 커질수록 모집단의 형태와 관계없이 정규분포에 근사하는 것을 볼 수 있으나 정규분포에 어느 정도 근사하는 위한 표본크기가 모집단의 형태에 따라 차이가 있음을 확인할 수 있음





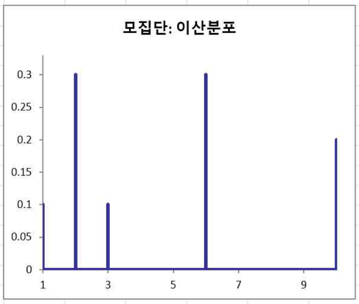
【분석결과 예제】

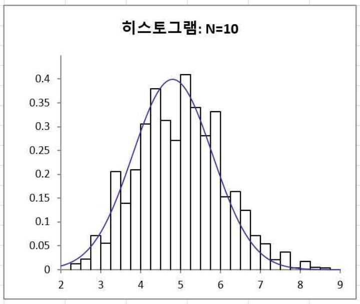
중심극한정리 모의실험: 이산분포

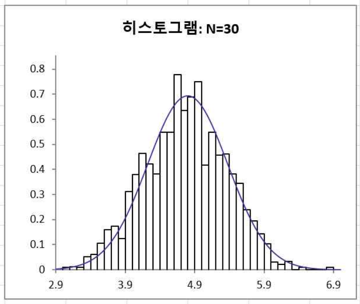
x	1	2	3	6	10
P(X=x)	0.100	0.300	0.100	0.300	0.200

모평균	4.800
모표준편ㅊ	3.156
반복수	10000

표준크기	표준오차	표본평균	표본표준편차
10	0.998	4.783	0.996
30	0.576	4.783	0.590







◉ 표집분포

○ 분포 및 추론 > 정규분포에서의 표집분포

- 정규분포에서의 확률표본을 이용하여 평균과 분산에 관련된 표집분포를 모의실험을 통해 유도하고 이론적인 분포와 비교함
- 중심축량선택에서

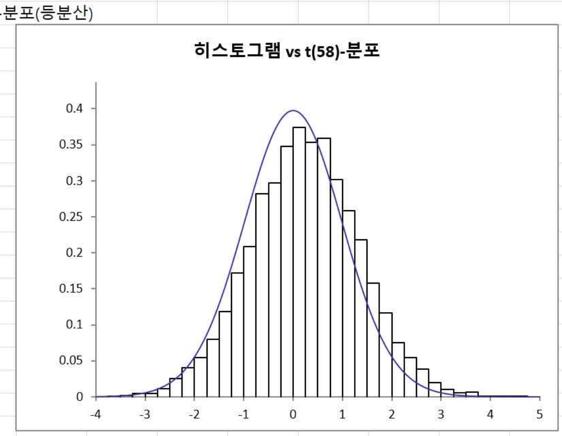
  - T-통계량(1표본): $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$
  - T-통계량(2표본: 등분산):
    $$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_p\sqrt{1/n_1+1/n_2}} \sim t_{n_1+n_2-2}$$
  - χ^2-통계량: $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$
  - F-통계량: $\dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}$

  모의실험을 통해 히스토그램을 작성하고 이론적인 분포와 비교함.
- T-통계량(2표본: 등분산)에서 분산이 다른데 같은 것으로 보고 위의 중심축량을 이용하여 모의실험을 해보면 그림과 같이 비대칭형태의 분포를 따르는 것을 확인할 수 있으며 이론적인 분포와 차이가 많이 있는 것을 볼 수 있음
- T-통계량(2표본: 이분산)의 경우 Welch–Satterthwaite equation을 통해 자유도를 계산하고 이 자유도는 표본분산의 값에 영향을 받기 때문에 이론적 분포를 유도할 수 없어 현재 버전에서는 제공하지 않음

【분석결과 예제】

정규분포 기반 표집분포 모의실험: t-분포(등분산)

가정	모집단1	모집단2
평균	0.000	0.000
표준편차	1.000	1.732
표본크기	30	30



◉ 신뢰구간

○ 분포 및 추론 > 정규분포에서의 신뢰구간

◦ 정규분포에서의 확률표본을 이용하여 평균과 분산에 관련된 신뢰구간을 모의실험을 통해 유도하여 신뢰수준과 포함확률을 비교함

◦ 추정모수에서

- 평균(1표본): $\overline{X} \pm t_{\alpha/2,n-1} S/\sqrt{n}$
- 평균차(2표본: 등분산):

$$\overline{X} - \overline{Y} \pm t_{\alpha/2, n_1+n_2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

- 평균차(2표본: 등분산):

$$\overline{X} - \overline{Y} \pm t_{\alpha/2, \nu} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

- 분산(1표본): $\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$, $\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}$
- 분산비: $\frac{S_1^2/S_2^2}{F_{1-\alpha/2,n_1-1,n_2-1}}$, $\frac{S_1^2/S_2^2}{F_{\alpha/2,n_1-1,n_2-1}}$

모의실험을 통해 히스토그램을 작성하고 이론적인 분포와 비교함.

◦ 평균차(2표본: 이분산)에서 자유도 ν는 Welch–Satterthwaite equation에 의해 유도되며 반복수는 최대 250까지 가능하며 결과에서 초과확률추정은 신뢰구간이 모수보다 위에 있는 비율, 미달확률추정은 신뢰구간이 모수 아래에 있는 비율이며 포함확률추정은 신뢰구간이 모수를 포함한 비율을 의미함

【분석결과 예제】

평균차에 대한 95% 신뢰구간

모평균	0	3
모표준편차	1.414	1.732
표본크기	30	30
반복수	100	
포함확률추정	0.970	
초과확률추정	0.020	
미달확률추정	0.010