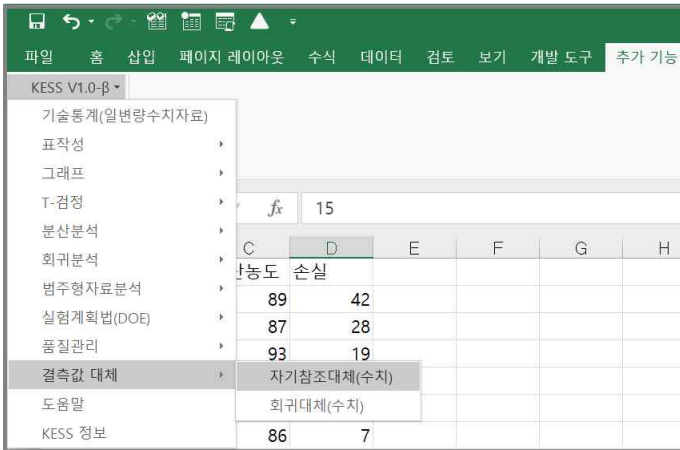


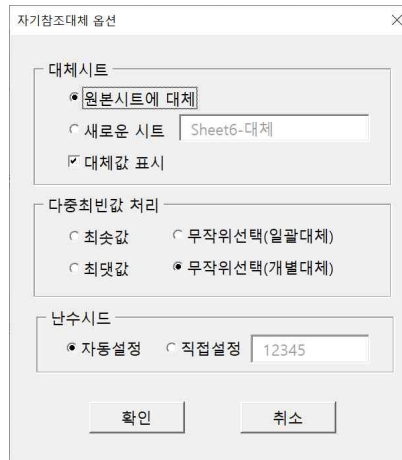
● 결측값 대체

○ 결측값 대체 > 자기참조대체(수치)

- 수량적 결측값을 평균, 중앙값, 최빈값, 최솟값, 최댓값으로 대체하거나 해당변수의 관측값을 무작위로 선택하여 대체함



- '변수목록'에서 대체할 변수를 선택해 '대체변수' 목록으로 전달.
- '대체방법'에서 대체값으로 사용할 통계값 선택. 'Hot Deck'은 대체값을 무작위로 선택함
- '대체옵션'을 누르면 '자기참조대체 옵션'이 나오고 대체시트명, 대체값표시(노란색), 최빈값이 여러 개인 경우 어떤 값으로 대체할지와 Hot Deck이나 다중최빈값 처리에서 난수시드를 지정할 수 있음



	A	B	C	D
1	나이	몸무게	클레스테롤	수축기혈압
2	38	51	160	110
3	28	60	188	
4	35	80	260	165
5	30	68.55556	210.8889	135
6	45	75	270	156
7	40	62	181	115
8	33	66	174	118
9	54	78	215	143
10	47	82	240	165
11	36	63	210	122

○ 결측값 대체 > 회귀대체(수치)

- 관측값들을 이용하여 회귀식을 유도하고 유도된 회귀식의 예측값 또는 예측변동성을 포함한 예측값으로 결측값을 대체함

KESS V1.0-β

기술통계(일반량수치자료)

표작성

그래프

T-검정

분산분석

회귀분석

범주형자료분석

실험계획법(DOE)

품질관리

결측값 대체

도움말

KESS 정보

fx 62

C	D	E	F	G	H
클레스테롤수축기혈압					
160	110				
188	125				
260	165				
	자기참조대체(수치)				
	회귀대체(수치)				
181	115				

회귀대체

변수목록

반응변수

수축기혈압

설명변수

나이

몸무게

클레스테롤

분석

출력옵션

재설정

도움말

종료

결측대체방법

☒ 예측기댓값대체
 ☐ 확률적대체(기댓값)
 ☐ 확률적대체(예측값)

- '변수목록'에서 '반응변수', '설명변수' 목록으로 전달
- 결측대체방법에서의 세 방법은 다음과 같음
 - 예측기댓값대체: 일반적으로 \hat{y} 로 표시되는 값으로 대체
 - 확률적 대체(기댓값): $\hat{y} + t \times se(\hat{y})$ 로 대체함. 여기서 t 는 t-분포를 따르는 난수
 - 확률적 대체(예측값): $\hat{y} + t \times se(y - \hat{y})$ 로 대체함.
- 출력옵션은 '자기참조대체 옵션'에서 '다중최빈값 처리' 부분이 빠진 것을 제외하고 동일함
- 설명변수가 없는 자료의 경우 예측값을 계산할 수 없기 때문에 대체되지 않음

	A	B	C	D
1	나이	몸무게	클레스테롤	수축기혈압
2	38	51	160	110
3	28	60	188	125
4	35	80	260	165
5	30	62	225	135
6	45	75	270	156
7	40	62	181	115
8	33	66	174	127.9017
9	54	78	215	143
10	47	82	240	165
11	36	63	210	122