웹R을 이용한 통계분석

문 건 웅

2020-05-12 15:10:56

머리말 - 웹R을 이용한 통계분석 책 pdf 출간

안녕하세요? 웹R(web-r.org)을 운영하는 문건웅입니다. "웹에서 하는 R통계"는 통계에는 관심이 있으나 R을 어려워하는 여러 연구자들을 위한 프로젝트입니다. R 설치없이 클릭만으로 웹에 있는 서버를 이용하여 통계분석을 하고 보다 R을 쉽게 사용하기 위한 패키지 개발 및 Shiny app 공동개발을 목표로 하고 있습니다. "웹에서 하는 R 통계분석" 책은 2015년 한나래출판사에서 출간되어 2쇄까지 완판되었으나 R이 업데이트되고 웹에서 하는 R 통계분석 앱의 내용이 계속 업데이트되면서 그 내용이 현재 프로그램을 제대로 반영하고 있지 못해 개정판을 내달라는 요구가 많았으나 기존과 같이 출판사를 통해 출간하는 방식으로는 웹에서 하는 R 통계분석 앱의 변화에 제대로 대응하지 못할 것으로 판단하여 출간을 미루어 왔습니다. 하지만 "웹R을 이용한 통계분석"책을 원하는 분들이 많아 leanpub(https://leanpub.com/webr) 을 통해 pdf형식으로 출간하였습니다. 아직 모든 내용이 책으로 정리되지는 않았으나 일단 leanpub 을 통하여 구매하실 수 있고 책을 구매하신 분들에게는 향후 내용이 추가되는 대로 이메일을 통해 알려드림으로써 다시 다운로드 받으실 수 있습니다. leanpub을 통해 책을 만들고 pdf 형태로 판매하는 경우 다음과 같은 장점이 있습니다.

독자입장에서의 장점

독자 입장에서는 이 책을 구입할 때의 장점은 크게 세 가지가 있습니다. 첫째, 책을 pdf형식으로 언제든지 구입하실 수 있으며 가격을 독자분이 정할 수 있다는 점입니다. 이 책을 구입하실 때에는 권장가격(suggested price)은 있으나 무료로도 구입하실 수 있습니다. 따라서 책을 다운로드 받아 읽어보신 후 값을 지불하실 수도 있습니다. 둘째, 한 번 구입한 책은 여러 번 다운로드 받으실 수 있습니다. 셋째, 책의 내용 중 일부가 개정되는 경우 이메일을 통해 알려주므로 언제라도 개정판을 다운로드 받으실 수 있습니다. 따라서 웹R을 이용한 통계분석처럼 내용이 자주 업데이트 되는 경우 항상 업데이트 된 내용의 책을 다운로드 받으실 수 있습니다. 그 외에 다른 사람들에게 책을 권할 때 책을 빌려주거나 pdf를 복사해줄 필요가 없습니다. 책을 다운로드 받을 수 있는 인터넷주소만 알려주면 됩니다.

저자입장에서의 장점

저자입장에서의 장점은 첫째, 내용의 업데이트가 자유롭다는 점입니다. 이 점은 독자입장에서의 장점이기도 하지만 저자로서도 매우 큰 장점이라고 할 수 있습니다. 책을 쓴다는 것은 책의 내용에 대해 책임을 져야 하는데 책에서 다루는 내용이 R과 같이 계속 업데이트 되는 경우 책을 쓰는 시점에서 잘 실행되던 코드가 R이 업데이트 되면서 에러가 발생할 수 도 있습니다. 이런 경우 독자 입장에서는 책의 내용에 불만을 토로할 수 있지만 저자 입장에서도 난감한 일입니다. 이제 인터넷을 통해 pdf로 출판을 하게 되면 수시로 책의 내용을 업데이트하더라도 다시 발간하는 비용이

들지 않고 독자들에게도 책의 바뀐 내용이 바로 전달될 수 있으므로 프로그램의 업데이트에 대한 부담없이 책을 쓸 수 있습니다. 둘째, 인세 문제인데 출판사를 통해 책을 출간하는 경우 저자가 받는 인세는 책값의 10 %정도에 지나지 않습니다. 물론 백만권씩 팔리는 밀리온셀러라면 10%가 큰 돈이 될 수 있지만 전문서적인 경우 100권이상 팔리는 전문서적이 많지 않습니다. 예를 들어 제 책의 정가가 4만원이고 천권이 팔렸다고 하면 저자가 받을 수 있는 인세는 약 400만원이지만 세금을 공제하고 나면 실제 받을 수 있는 인세는 약 9% 정도입니다. 책을 쓰는데 들어간 시간과 노력을 생각하면 정말 적은 금액입니다. leanpub을 통해 출간하면 저자가 인세를 80% 받게 됩니다. 물론 pdf 출간이며 책 가격을 독자가 정할 수 있고 얼마든지 복사하여 유통할 수 있으므로 실제로 저자가 받게 되는 인세는 많지 않겠지만 그럼에도 불구하고 책을 정가로 구입하고 불법복사하지 않는다는 출판문화의 정착을 위해서도 pdf 출간은 바람직한 출판형태라고 판단됩니다. 물론 종이책을 원하시는 분은 종이책을 구입하실 수도 있습니다.

위와 같은 장점들이 있어 저는 국내의 여러 출판사들에게 이와 같은 형태의 pdf출간을 의뢰하여 보았지만 아직 국내 출판사의 입장에서는 pdf 출판을 하기 어려운 것 같습니다. 저는 위에서 열거한 여러 장점 중 업데이트가 자유롭다는 점이 가장 마음에 듭니다. 여러분들께서는 이 책을 구입해주심으로써 웹에서 하는 R 통계를 후원해주시는 것이며 웹에서 하는 R통계를 지속적으로 발전시켜 나갈 수 있는 힘과 용기를 주시는 것입니다. 또한 pdf 책을 구입해주시는 분들은 bookdown으로 만든 웹R책의 온라인버젼을 자유롭게 보실 수 있습니다. 웹R을 이용한 통계분석 책은 2020년 5월 12일 전처리하기를 시작으로 표만들기, 기술통계, 탐색적그래프, 비교통계 순으로 업데이트 예정입니다. leanpub을 통해 책을 구입하신 분들은 책이 업데이트 될 때마다 안내 이메일을 받으실 수 있습니다. 감사합니다.

2020년 5월 12일

문건웅

차 례

I 웹R을 이용한 데이터 전처리 1

제 1 장 데이터 전처리 맛보기 3

제 1 절 웹R 접속하기 . . . 3
제 2 절 로그인 및 무료서버 접속하기 . . . 4
제 3 절 웹에서 하는 R통계분석 첫 화면 . . . 4
제 4 절 데이터 전처리 예제 파일 다운로드 받기 . . . 5
제 5 절 예제 데이터 업로드 . . . 6
제 6 절 데이터 전처리 예제 선택 . . . 7
제 7 절 데이터 전처리 . . . 7
7.1 열이름 정리하기 . . . 8
7.2 되돌리기와 다운로드 . . . 8
7.3 비어 있는 행/열 삭제 . . . 9
7.4 엑셀숫자를 날짜로 . . . 10
7.5 문자열을 날짜로 . . . 10
7.6 결측치(NA)로 만들기 . . . 11
7.7 열 합병하기 . . . 12
7.8 중복 데이터 찾기 . . . 13

제 2 장 간단한 설문조사 데이터 전처리 15

제 1 절 데이터 업로드 . . . 16
제 2 절 데이터 정리 . . . 16
제 3 절 데이터 업로드 . . . 17
제 4 절 첫번째 행을 라벨로 사용 . . . 17
제 5 절 라벨붙이기 - 담당 . . . 18
제 6 절 라벨붙이기 - 근무기간 . . . 19
제 7 절 한꺼번에 라벨 붙이기 - 내용평가 및 교육만족도 . . . 19

제 8 절 결측치의 처리 21
제 9 절 라벨붙이기 22
제 10 절 역순으로 만들기 22
제 11 절 합계 및 평균 구하기 23
제 12 절 전처리 끝난 자료 다운로드 23

제 3 장 실전 설문조사 데이터 전처리 25
제 1 절 데이터 불러오기 26
제 2 절 일치도(Cronbach의 alpha) 26
제 3 절 자동으로 코드 역순으로 28
제 4 절 합계구하기(1) 29
제 5 절 자동역순처리합계 30
제 6 절 일치도 구하기 : 간호사의 태도 30
제 7 절 평균중심화와 표준화하기 33
제 8 절 0/1로 입력되어 있는 자료를 바꾸기 35
제 9 절 범주형 변수의 순서 정하기 36
제 10 절 계산해서 새로운 열 만들기 37
제 11 절 서브그룹 만들기1 38
제 12 절 서브그룹 만들기2 39
제 13 절 긴 형태로 바꾸기(pivot_longer) 40
제 14 절 넓은 형태로 바꾸기(pivot_wider) 42
제 15 절 두 개의 데이터 병합하기 43
제 16 절 전처리가 끝난 데이터 다운로드 44

편 I

웹R을 이용한 데이터 전처리

제 1 장

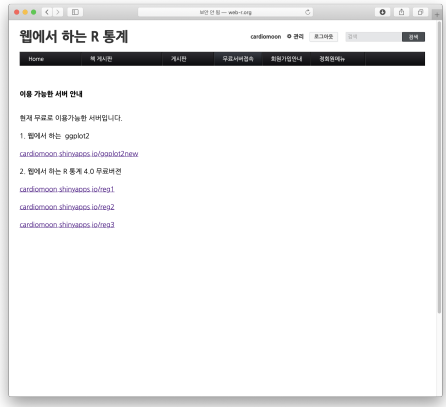
데이터 전처리 맛보기

제 1 절 웹R 접속하기

인터넷 브라우져를 통해 웹에서 하는 R 통계분석(web-r.org)에 접속한다. 웹에서 하는 R통계분석은 R 과 shiny를 이용하여 만든 shiny app으로 이를 이용하기 위해서는 HTML5를 지원하는 웹브라우져가 필요하다. 구글 chrome이나 safari 등의 웹브라우져를 이용하여 접속할 것을 권장한다. 인터넷 익스플로러에서는 테스트해보지 않아 제대로 동작하는지 보장하지 못한다.

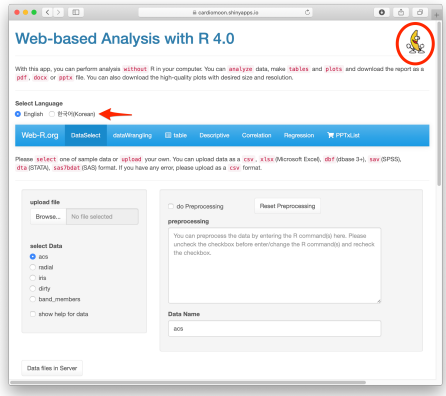
제 2 절　로그인 및 무료서버 접속하기

웹R의 가입은 무료이므로 가입 후 로그인을 하면 “무료서버접속” 메뉴에서 무료서버 접속을 할 수 있다. 웹에서 하는 R 통계분석 4.0 무료버젼 중 하나를 선택해 클릭한다.



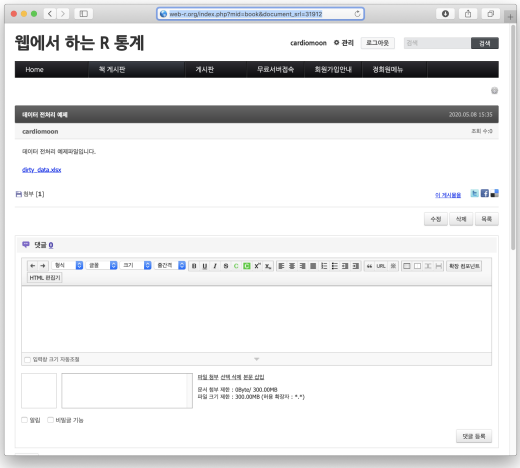
제 3 절　웹에서 하는 R통계분석 첫 화면

잠시 기다리면 다음과 같은 화면이 나타난다. 처음 로딩할 때는 약 30초-1분 가량 걸리나 이후에는 프로그램의 로딩에 10초 정도 걸린다. 동그라미 부분에 보면 춤추는 바나나인형이 보이는데 이 인형이 춤추고 있는 경우 프로그램 내부에서 계산이 진행되고 있는 것이므로 춤을 멈출 때까지 기다렸다가 진행하는 것이 바람직하다. 영어가 편한 분들은 이 상태에서 진행하여도 무방하지만 한국어가 익숙하신 분들은 select languge에서 한국어를 선택한다(화살표).

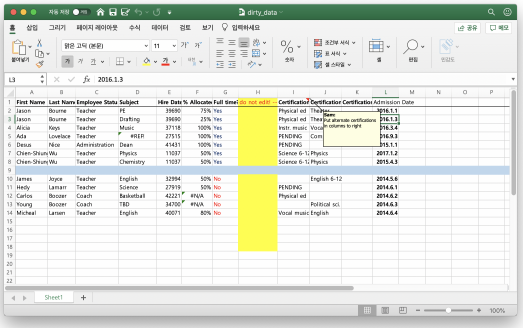


제 4 절 데이터 전처리 예제 파일 다운로드 받기

다음 주소에서 데이터 전처리에 쓰일 파일을 다운로드 받는다. 다운로드 받을 파일은 dirty_data.xlsx 로 http://web-r.org/index.php?mid=book&document_srl=31912 에서 다운로드 받을 수 있다.



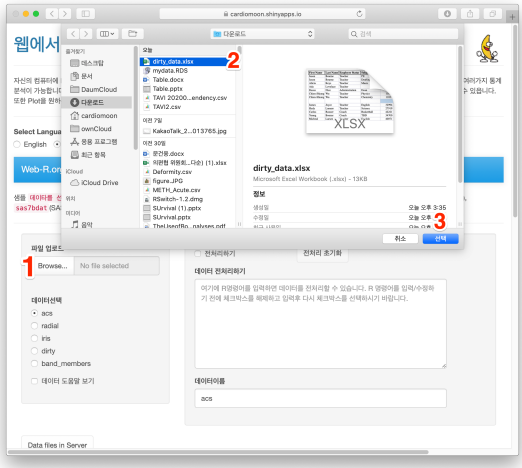
다운로드 받은 파일을 엑셀로 열어보면 다음과 같다.



참고로 R에서는 열이름에 공백이 있으면 안되고 %등 특수문자나 숫자로 열이름이 시작해서는 안된다. 엑셀에서는 상관이 없다.

제 5 절 예제 데이터 업로드

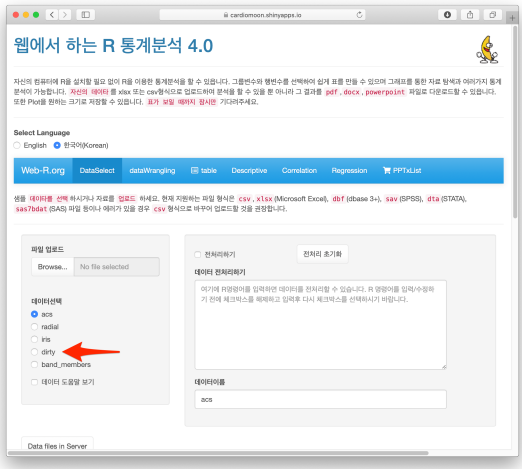
위에서 다운로드 받은 파일을 업로드하기 위해 Browse... 버튼(1)을 누르고 다운로드 받은 파일을 선택한 후(2) 선택버튼을 누른다(3).



업로드가 되면 “upload complete”메시지가 나타나고(화살표) 데이터이름이 uploaded로 바뀐다(4).

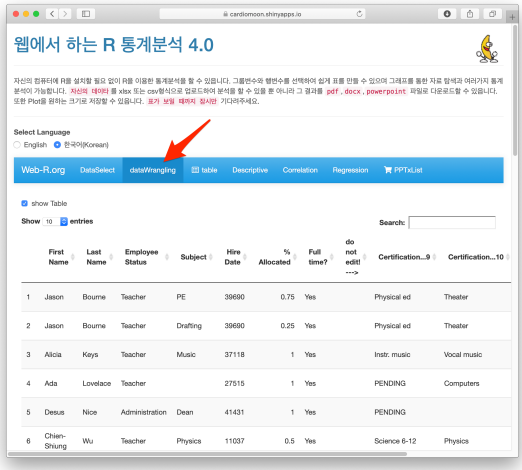
제 6 절 데이터 전처리 예제 선택

2.4 - 2.5의 다운로드와 업로드 과정 없이 데이터선택에서 dirty를 선택해도 같은 결과가 나타난다.



제 7 절 데이터 전처리

이 데이터를 바로 사용하여 분석을 진행하면 에러가 난다. 먼저 메뉴에서 dataWrangling을 선택한다.



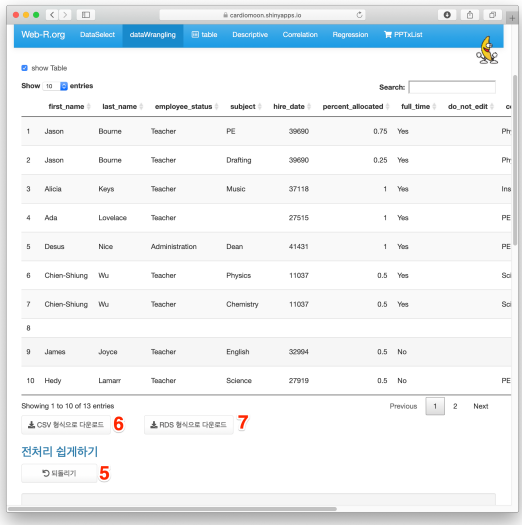
7.1 열이름 정리하기

화면을 아래로 이동하여 전처리선택하기에서 “열이름 정리하기”가 선택되어 있는 것을 확인하고 *열이름정리하기* 버튼을 누른다. R에서는 열이름에 공백문자를 사용할 수 없고 특수문자도 사용할 수 없다. *열이름정리하기* 는 열 이름을 R에서 사용할 수 있도록 바꾸어 준다.

다시 화면을 위로 올려보면 열이름 중 공백이 _글자로 대치되고(예를 들어 first name이 first_name) %는 percent로 바뀌어 있는 것을 알수 있다. 또한 certification이라는 열 이름이 중복이 되어 있었는데 certification_9, certification_10으로 바뀌어 있는 것을 확인할 수 있다.

7.2 되돌리기와 다운로드

전처리 도중 전처리를 잘못한 경우를 대비하여 되돌리기 버튼(5)이 있는데 이 버튼을 누르면 바로 전의 상태로 되돌아간다. 또한 전처리가 끝난 데이터를 csv형식으로 다운로드(6) 받거나 RDS형식으로 다운로드(7) 받을 수 있는데 여러가지 이유에서 RDS형식으로 다운로드 받을 것을 권한다.

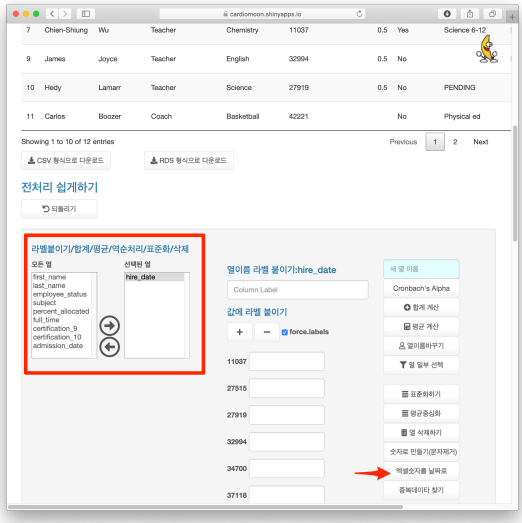


7.3 비어 있는 행/열 삭제

데이터를 보면 8번째 행이 비어있고 8번째 열 또한 비어있는 것을 확인할 수 있다. 경우에 따라 엑셀 파일을 업로드한 경우 데이터의 아래에 빈 행이 여러 개 나타나는 경우도 있다. 이런 경우 제대로 분석이 되지 않으므로 비어있는 행/열을 삭제할 것을 권한다. 화면을 아래로 이동하여 전처리 선택하기 중 “비어있는 행/열 삭제”를 선택한 후(사각형) “비어있는 행/열 삭제” 버튼을 누른다(화살표).

7.4 엑셀숫자를 날짜로

엑셀에서 날짜를 입력한 경우 R로 불러오면 숫자로 변환되어 있는 경우가 있다. 이 데이터의 경우 hire_date가 숫자로 변환되어 있는데 이를 날짜로 바꾸려면 먼저 모든 열에서 hire_date 열을 선택하여 오른쪽으로 옮긴후 "엑셀숫자를 날짜로" 버튼을 누르면 된다.



7.5 문자열을 날짜로

위의 데이터 중 admission_date는 문자로 변환되어 있다. 이를 날짜로 바꾸려면 먼저 admission_date를 선택한 후(1) 전처리 선택하기에서 "문자열을 날짜로" 를 선택한 후(2) "날짜로 바꾸기" 버튼을 누른다(화살표).

7.6 결측치(NA)로 만들기

경우에 따라 데이터가 누락되어 있는 경우 99, 999, NA등으로 표시하는 경우가 있다. 이 데이터에서 PENDING으로 표시되어 있는 것을 결측치로 만들려면 “NA(결측치)로 만들기”를 선택한 후(1) PENDING을 입력하고(2) “결측치(NA)로 만들기” 버튼을 누른다(화살표). 이 때 선택된 열(들)이 있는 경우 그 열(들)만 결측치 만들기가 진행되며 선택된 열이 없는 경우(3) 모든 열에 대해 변환이 진행된다.

7.7 열 합병하기

경우에 따라 두 개의 열에 있는 데이터를 병합하여 새로운 열을 만들어야 할 때가 있다. 이 데이터에서 certification_9와 certification_10에 있는 데이터 중 누락되지 않은 첫번째 값을 선택하려면 먼저 certification_9와 certification_10 두 개의 열을 선택하고(1) 전처리 선택하기 중 합병하기를 선택한 후(2) 합병하기 버튼을 누른다 (화살표). 데이터의 마지막에 certification_All이라는 열에 두 개의 열이 합병되어 나타난다.

7.8 중복 데이터 찾기

데이터 중 중복된 값을 찾는 것ㅇ은 매우 간단하다. first_name과 last_name이 같은 데이터를 찾으려면 first_name과 last_name을 선택한 후(1) 중복데이터 찾기를 누르면 된다(화살표). 중복데이터를 찾은후 되돌리기를 누르면 다시 전의 상태로 돌아간다.

다음 장에서는 실제 설문조사 데이터를 가지고 전처리를 연습해본다.

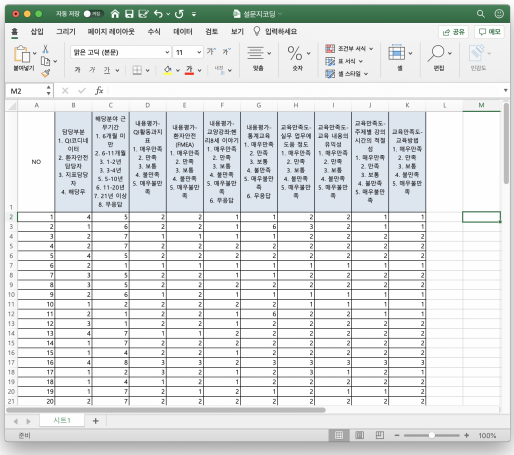
제 2 장

간단한 설문조사 데이터 전처리

이번 장에서는 간단한 실제 설문조사 데이터를 가지고 데이터 전처리를 해본다. 사용할 데이터는 다음 주소에서 다운로드 받을 수 있다(http://web-r.org/index.php?mid=book&document$_{s}$rl=31924). 이 데이터는 성빈센트병원에서 제 5회 QI(Quality Improvement) Academy 가 끝난 후 시행한 평가 설문지 데이터로 설문 내용은 “설문지.docx”이다. 먼저 설문지코딩.xlsx 파일을 살펴보자.

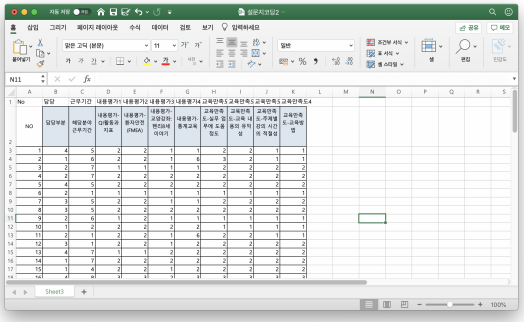
제 1 절　데이터 업로드

“설문지코딩.xlsx” 파일을 엑셀로 열어보면 다음과 같다. 첫번째 줄에 설문 제목과 내용이 모두 정리 되어 있다. 이 파일을 그대로 R에서 불러 사용하면 열이름이 너무 길어 데이터를 분석할 때 힘들어진다. 또한 입력단계에서는 각 항목 라벨이 의미있으나 열이름에 라벨이 포함될 필요는 없다.



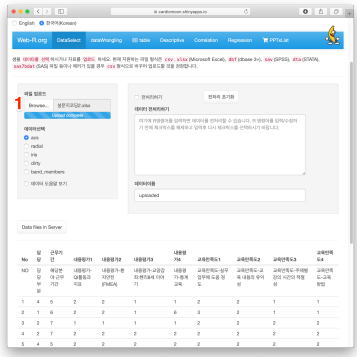
제 2 절　데이터 정리

가장 권하고 싶은 방법은 이와 같이 정리를 하는 것이다.열 이름은 짧을수록 좋으며 빈칸이 있으면 안된다. 또한 숫자로 시작하지 않는 것이 좋다. 단어 사이를 띄어쓰기를 하고 싶은 경우 언더바를 사용하는 것이 좋다. (예: 언더_바) 먼저 열이름을 최대한 간단하게 이름 붙이고 두번째 행에 자세한 내용을 적는다. 두번째 행의 내용은 이후 웹R에서 열이름 라벨로 사용된다. 이 자료파일의 이름은 설문지코딩2.xlsx이다.



제 3 절 데이터 업로드

웹R에 접속한 후 무료서버접속 페이지를 통해 무료서버에 접속한다. 파일 업로드를 위해 Browse... 버튼을 누르고(1) 다운로드 받은 "설문지코딩2.xlsx" 파일을 업로드 한다.



제 4 절 첫번째 행을 라벨로 사용

메인 메뉴에서 dataWrangling을 선택한다.

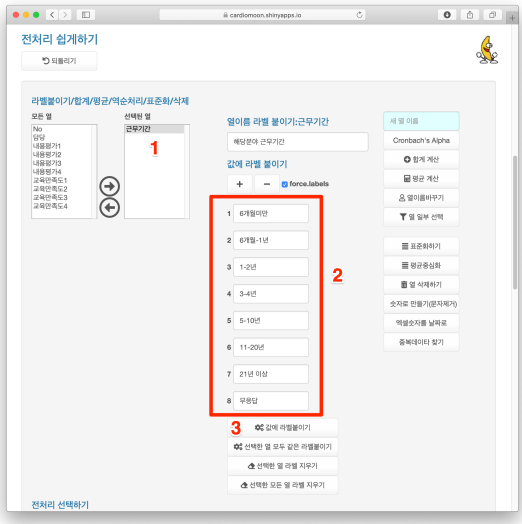
화면을 아래쪽으로 이동하여 "전처리 선택하기" 중 "첫번째 행을 라벨로 사용"을 선택하고(1) "첫번째 행을 라벨로 사용"버튼을 누른다(화살표).

제 5 절 라벨붙이기 - 담당

라벨붙이기/합계/평균/역순처리/표준화/삭제에서 담당 열을 선택한 후(1) 라벨을 입력하고(2) **값에 라벨붙이기** 버튼(화살표)을 누른다.

제 6 절 라벨붙이기 - 근무기간

근무기간을 선택하고(1) 1-8까지 해당하는 라벨을 입력한 후(2) **값에 라벨붙이기 버튼**을 누른다(3).



제 7 절 한꺼번에 라벨 붙이기 - 내용평가 및 교육만족도

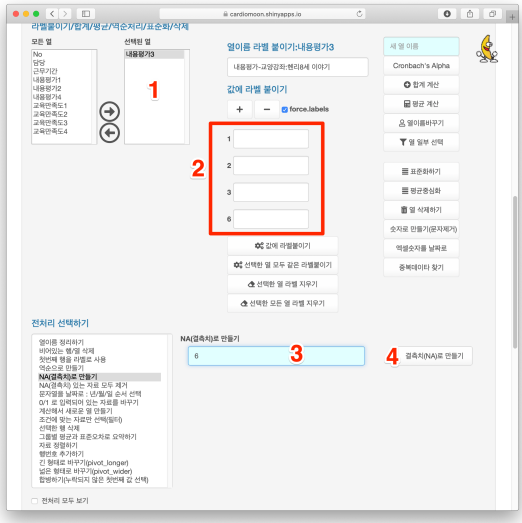
내용평가1-4, 교육만족도1-4를 선택한 후 교육만족도4 하나만 선택해보면(1) 1-4까지 빈칸이 보인다. 이 항목들은 5단계 리커트 척도를 사용하였는데 리커트 척도를 사용하는 경우 응답자는 극단적인 선택을 피하려는 경향이 있기 때문에 매우 불만족을 선택한 사람이 아무도 없는 것을 알 수 있다. 모두 5단계가 있으므로 + 버튼(2)을 누른다.

계속해서 1-5단계까지 라벨을 입력하고(1) 선택한 열 모두 같은 라벨붙이기 버튼을 누른다(2).

다음과 같은 경고메시지가 나타난다. 내용평가3은 한명이 값을 입력하지 않아 6으로 코딩한 값이 있어 이와 같은 경고메시지가 나타난다. 당황하지 말고 닫기버튼을 누른다.

제 8 절 결측치의 처리

내용평가3을 선택하면(1) 1,2,3,6 이 있는 것이 보인다(2). 이중 6은 결측치이다. 화면 아래쪽에 있는 NA(결측치)로 만들기를 선택하고 6을 입력하고(3) 결측치(NA)로 만들기 버튼을 누른다(4).

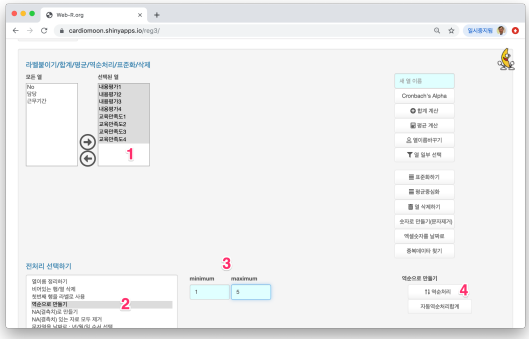


제 9 절 라벨붙이기

내용평가3에 라벨이 없으므로 내용평가3과 교육만족도4를 선택하고(1) 선택한 열 모두 같은 라벨붙이기 버튼을 누른다(2).

제 10 절 역순으로 만들기

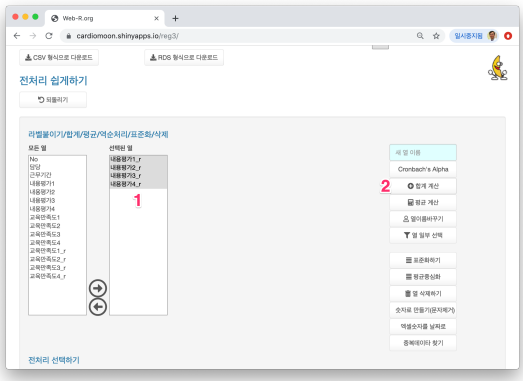
내용평가1-4 와 교육만족도1-4는 모두 매우 만족이 1, 매우 불만족이 5로 코딩되어 있다. 이를 뒤집어 매우 만족이 5, 매우 불만족이 1로 바꾼다면 점수가 높을수록 만족도가 높다는 것을 알수 있을 것이다. 이렇게 항목을 역순으로 만들려면 다음과 같이 한다. 먼저 내용평가1-4, 교육만족도1-4를 모두 선택하고(1) 전처리선택하기 중 역순으로 만들기를 선택한 후(2) minimum에 1, maximum에 5를 입력하고(3) **역순처리** 버튼을 누른다(4).



역순처리가 끝나면 열이름 끝에 _r이 붙은 열이 새로 생긴다. 즉 **내용평가1** 열의 역순처리한 열은 **내용평가1_r** 열이다.

제 11 절 합계 및 평균 구하기

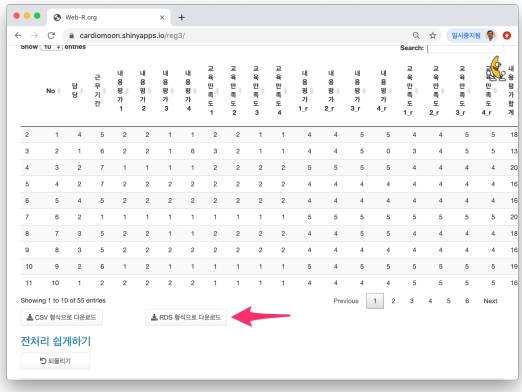
내용평가 및 교육만족도의 합계 및 평균을 구해보자. 먼저 내용평가1_r 부터 내용평가4_r까지 네개의 열을 선택한 후 합계계산 버튼을 누른다(3). 내용평가합계 열이 새로 만들어지며 여기에 합계가 기록된다.



계속해서 내용평가1_r 부터 내용평가4_r까지 네개의 열을 선택한 후 평균계산 버튼을 누른다. 교육만족도의 합계와 평균 또한 같은 방법으로 구한다.

제 12 절 전처리 끝난 자료 다운로드

위와 같은 과정을 거쳐 전처리를 한 후 전처리된 자료를 다운로드하여 저장하면 다음에 이 자료를 이용하여 분석할 때 다시 전처리를 할 필요가 없다. 화면을 위로 이동한 후 **RDS 형식으로 다운로드** 버튼을 누른다. RDS형식으로 저장하면 전처리 과정에서 붙인 라벨 등이 그대로 유지된다. csv형식으로 저장하는 경우 라벨등이 없어지며 날짜 형식의 자료도 문자형으로 바뀌니 주의하여야 한다.



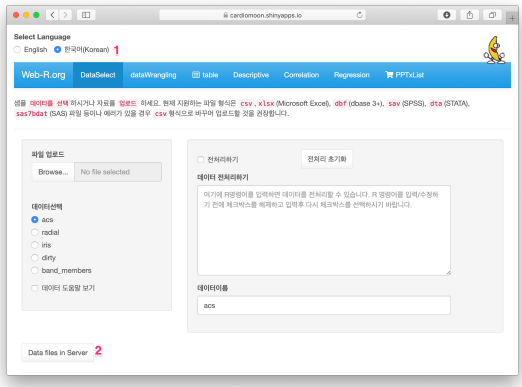
제 3 장

실전 설문조사 데이터 전처리

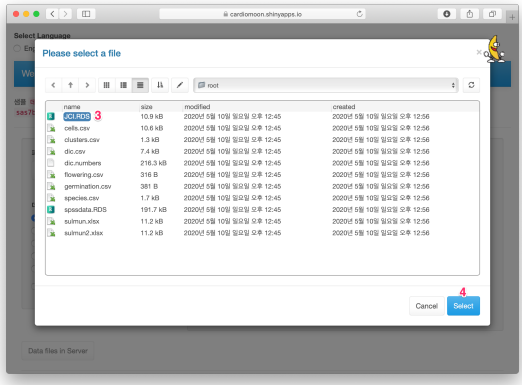
이번 장에서는 JCI 인증에 대한 임상간호사의 인식과 태도, 직무스트레스라는 연구에 사용된 테이터를 가지고 전처리 연습을 해본다. 사용할 데이터(JCI인증.xlsx) 및 논문(JCI인증.hwp)은 다음 주소에서 다운로드 받을 수 있다(http://web-r.org/book/14125). 자료를 다운받지 않아도 웹R의 무료서버에 접속하여 내장된 데이터를 이용할 수 있다.

제 1 절 데이터 불러오기

웹R에 접속한 후 무료서버접속 페이지를 통해 무료서버에 접속한다. 먼저 언어를 한국어로 선택한다. 서버에 내장되어 있는 데이터를 불러오기 위해 **Data files in Sever** 버튼을 누른다(2).



내장되어 있는 데이터 중 JCI.RDS 파일을 선택하고(3) Select 버튼을 누른다(4).



제 2 절 일치도(Cronbach의 alpha)

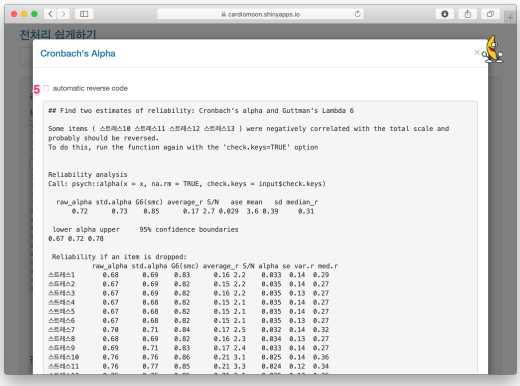
이 데이타는 JCI 인증에 대한 임상간호사의 인식과 태도, 직무스트레스에 관한 연구로 직무스트레스에 관한 설문은 모두 13개로 다음과 같다.

1. JCI 인증을 준비하면서 추가업무가 많아 항상 시간에 쫓기며 일했다.
2. JCI 인증을 준비하면서 나는 직무로 인해 잠을 깊이 자지 못하고 자주 깼다.
3. JCI 인증을 준비하면서 평가업무가 힘들어 이직을 고려해 본적이 있다.

4. JCI 인증을 준비하면서 행정적 업무로 인해 연장근무를 하였다.
5. JCI 인증을 준비하면서 직접적인 현장활동 업무강화로 연장근무를 하였다.
6. JCI 인증을 준비하면서 나의 심적 스트레스가 증가하였다.
7. 우리병원은 평가 업무 분장시 부서간 마찰로 업무협조가 잘 이루어지지 않았다.
8. JCI 인증을 준비하면서 내부구성원의 불만이 증가 하였다.
9. JCI 인증을 준비하면서 부가적 업무준비로 고객접점 관리가 소홀하였다.
10. JCI 인증 준비로 인한 업무표준화로 고객접점 관리에 효과적으로 대응을 할 수 있었다.
11. 나의 능력을 개발하고 발휘할 수 있는 기회가 주어진다.
12. 나의 업무수행 과정에서 나는 결정할 권한이 주어지며 영향력을 행사 할 수 있다.
13. 일에 대한 나의 업적을 고려할 때 나는 직장에서 제대로 존중과 신임을 받고 있다.

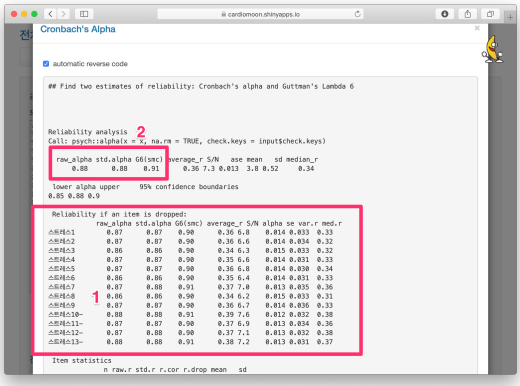
이들 설문을 자세히 보면 1-9까지는 설문의 내용이 JCI인증에 대한 부정적인 평가를 묻는 질문으로 되어 있고 10-13까지는 설문의 내용이 JCI인증에 대한 긍정적인 평가를 묻는 질문으로 되어 있다. 이들 설문에 대한 일치도를 계산하려면 다음과 같이 한다. 먼저 메인 메뉴의 dataWranging을 선택한 후 스트레스1-스트레스13까지의 13개의 열을 선택하고(1) Cronbach's alpha 버튼을 누른다(2).

다음 화면이 나타난다. 화면의 글을 읽어보면 일부의 항목(스트레스 10-13)이 전체 값과 음의 상관관계를 보이므로 코딩을 뒤집어 할 것을 권유하고 있다. 현 상태의 Cronbach의 alpha값은 0.73으로 나타난다. 프로그램에서 권하는 대로 automatic reverse code를 선택하면(5) 화면이 바뀐다.



제 3 절 자동으로 코드 역순으로

automatic reverse code를 선택하면 일부항목에서 스트레스10부터 스트레스13까지 스트레스10- 과 같이 항목에 -가 표시된다(1). 이 항목들은 자동으로 코드를 역순으로 취한 후 Cronbach의 alpha값을 계산한 것으로 이 때 alpha값은 0.73에서 0.88로 증가되는 것을 알 수 있다.



제 4 절 합계구하기(1)

간호사의 직무 스트레스의 합계와 평균을 구하려면 먼저 스트레스10부터 스트레스13까지의 항목을 역순으로 만든다. 먼저 스트레스10 - 스트레스13 항목을 선택한 후(1) 전처리 중 역순으로 만들기를 선택하고(2) 최소값에 1, 최대값에 5를 입력한 후(3) 역순처리 버튼을 누른다(4).

계속해서 스트레스1 - 스트레스9와 스트레스10_r부터 스트레스 13_r을 선택한 후(1) 합계 계산 또는 평균 계산 버튼을 눌러 계산한다.

제 5 절 자동역순처리합계

매번 위와 같이 Cronbach의 alpha값을 구하고 역순처리할 항목을 알아낸 후 역순처리하고 합계 및 평균을 계산하려면 무척 번거로울 뿐만 아니라 실수의 여지도 있다. 웹R에서는 이 과정을 자동으로 처리해준다. 먼저 스트레스1-13까지 항목을 선택한 후 전처리 중 역순으로 만들기를 선택하고(2) 최소값에 1, 최대값에 5를 입력한 후(3) 자동역순처리합계 버튼을 누른다(4). 웹R에서는 자동으로 역순처리할 항목들을 역순처리한 후 합계와 평균을 계산해준다. 즉 스트레스10_r 부터 스트레스13_r 과 스트레스합계, 스트레스평균이 계산된다.

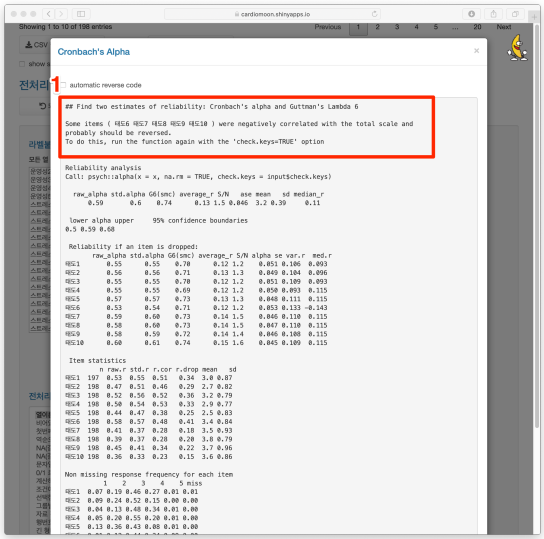
제 6 절 일치도 구하기 : 간호사의 태도

JCI 인증에 대한 간호사의 태도는 태도1 - 태도10에 정리되어 있다. 이 문항들을 살펴보면 다음과 같다. 1. 평가기준에 따른 적정한 시설, 장비, 인력 등을 확보할 수 있는 근거가 되었다. 2. 현재 갖추어진 지침에 따라 체계적으로 정착화 될 때까지 JCI인증을 받아야 한다. 3. 평가수검 우수 결과 도출로 대외적인 인지도 상승효과를 가져왔다. 4. 지속적인 평가 수검으로 우리병원 규모에 맞는 질적 경쟁력을 확보한다. 5. 현재 논의되고 있듯이 평가수검 병원에 대한 상대적인 보상체계가 주어졌다. 6. 공공병원 현실에 맞지 않는 평가기준이 많아 별도 평가기준으로 받아야한다. 7. 평가시 임시적인 대응이 많아 근본적인 질향상 효과가 없다. 8. 얻어지는 결과에 비해 투입될 비용부담이 크다. 9. 평가로 인한 부가적인 업무로 직접적인 고객응대에 더 소홀하게 된다. 10. 현재 병원의 평가시스템 미비로 평가대비 구조적 문제가 많다.

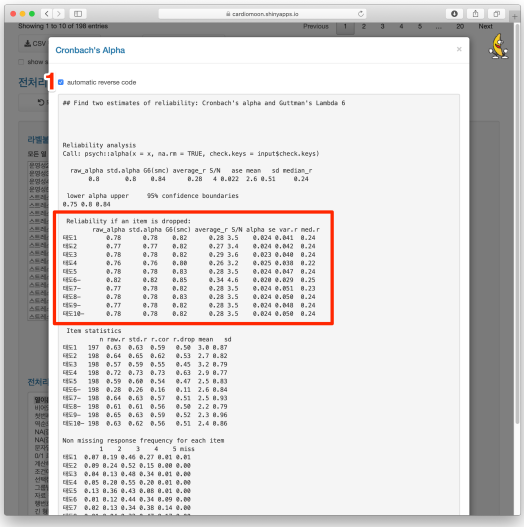
이들 문항들을 자세히 보면 1-5까지는 설문의 내용이 JCI인증에 대한 긍정적인 평가를 묻는 질문으로 되어 있고 6-10까지는 부정적인 평가를 묻는 질문으로 되어 있다.

이들 문항에 대한 일치도를 구하려면 태도1 - 태도10 을 선택한 후(1) Cronbach's Alpha 버튼을 누른다(2).

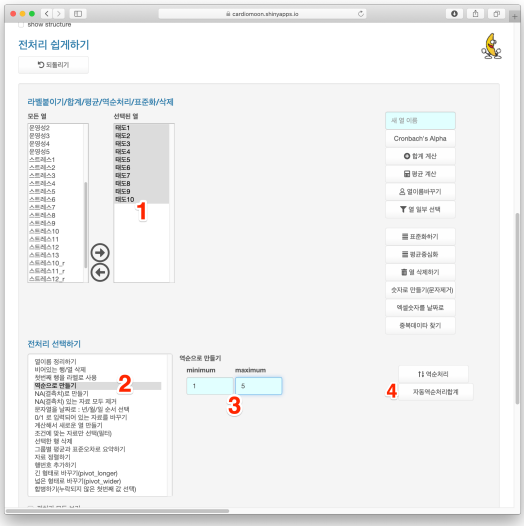
automatic reverse code 를 선택하지 않았을 때(1) 일치도(std.alpha) 는 0.6이며 태도6 부터 태도10까지는 전체 값과 비교할 때 음의 상관관계가 있으므로 코드를 역순처리할 것을 권하고 있다.



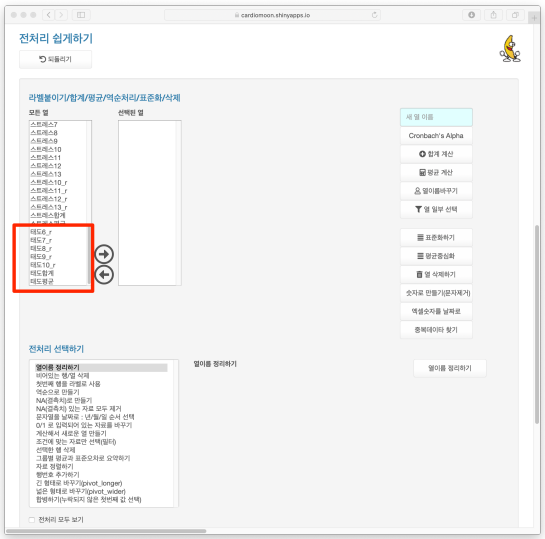
automatic reverse code 를 선택하면(1) 일치도(std.alpha) 가 0.8로 증가된다. 이때 태도6 부터 태도10까지는 태도6-과 같이 항목에 -가 표시된다.



계속해서 태도1 부터 태도10까지의 열이 선택되어 있는 상태(1)에서 전처리 중 역순으로 만들기(2)를 선택한 후 최소값에 1, 최대값에 5를 입력하고(3) 자동역순처리합계 버튼을 누른다(4).



태도6_r 부터 태도10_r 까지가 계산되어 있고 태도합계, 태도평균도 계산되어 있는 것을 알 수 있다.



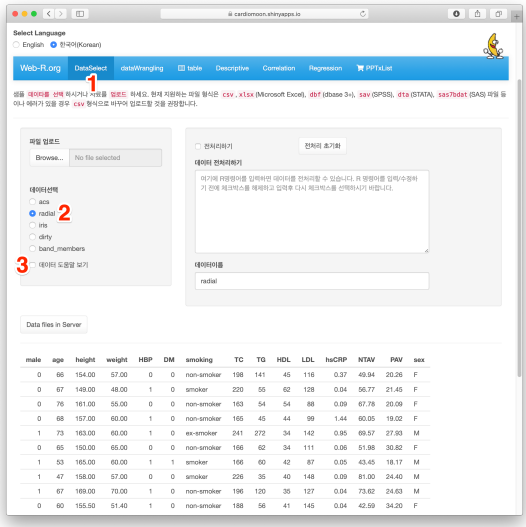
제 7 절 평균중심화와 표준화하기

연속형변수인 경우 평균중심화 또는 표준화를 시행할 경우가 있다. 평균중심화는 개별변수의 값에서 그 변수의 평균값을 뺀 것을 말한다. 또한 표준화는 개별변수의 값에서 그 변수의 평균값을 뺀 후 그 변수의 표준편차로 나눈 것을 말한다.

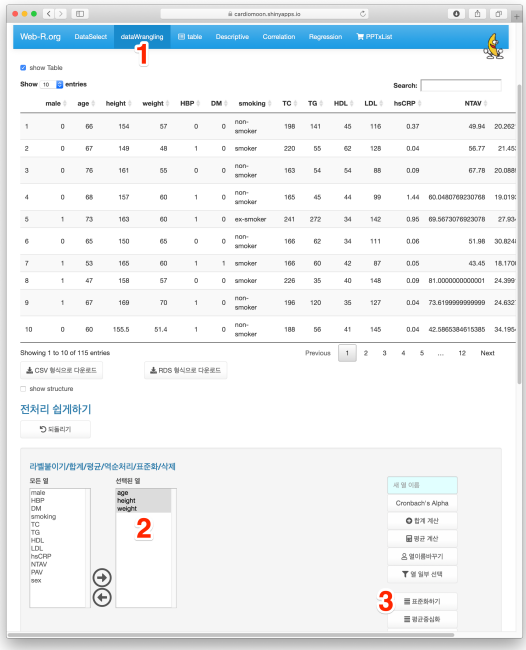
$$mean\ centering : x_i - \bar{x} \tag{3.1}$$

$$standardize : \frac{x_i - \bar{x}}{sd(x)} \tag{3.2}$$

먼저 메인 메뉴의 DataSelect(1)에서 두번째 예제인 radial(2)을 선택해 본다. 이 데이터에 대한 도움말을 보고 싶을 때에는 데이터 도움말 보기를 선택하면 된다(3).



이 데이터 중 환자의 나이와 키, 몸무게를 표준화해보자. 메인메뉴의 dataWrangling을 선택하고(1) age, height, weight의 세개의 열을 선택한 후(2) 표준화하기 버튼을 누른다(3).



표준화를 한 경우 열이름에 _std가 붙는다. 이 경우 표준화한 age_std, height_std 및 weight_std 열이 새로 생긴다. 평균중심화한 경우 _mc가 붙는다.

제 8 절 0/1로 입력되어 있는 자료를 바꾸기

radial 데이터에는 고혈압(HBP), 당뇨(DM) 등의 병력이 없는 경우는 0, 있는 경우는 1로 입력되어 있다. 이 경우 라벨붙이기를 통해 0, 1 값을 고혈압유무로 라벨을 붙일 수 도 있고 아예 0/1의 값을 HBP(-)/HBP(+) 와 같이 바꿀 수도 있다. HBP를선택한 후(1) 전처리 중 0/1로 입력되어 있는 자료를 바꾸기(2)를 선택하고 열이름포함 및 (-)/(+)를 선택한 후(3) **0/1 다시코딩** 버튼(4)을 누르면 HBP 열의 0과 1이 각각 HBP(-), HBP(+) 로 바뀌는 것을 알 수 있다.

제 9 절 범주형 변수의 순서 정하기

radial 데이터의 smoking은 흡연여부가 기록되어 있는데 ex-smoker, non-smoker, smoker의 세 개의 값으로 되어 있다. 이 경우 따로 순서를 정하지 않는 경우 알파벳 순서에 의해 ex-smoker, non-smoker, 순으로 통계처리가 되지만 연구자에 따라 이 세 개의 값 중 비흡연(non-smoking), 담배를 끊은 사람(ex-smoker), 현재 흡연하는 사람(smoker)의 순서로 처리되는 것을 원할 수 있다. 이 경우 범주형변수의 순서를 정해주어야 하는데 먼저 smoking을 선택한 후(1) 범주형변수의 순서정하기에서 non-smoker, ex-smoker, smoker 순으로 클릭한 후(2) **Factor로 바꾸기** 버튼(3)을 누르면 순서가 바뀐다. 이 순서는 show structure(4) 를 선택해 데이터의 구조를 보면 확인할 수 있다.

제 10 절 계산해서 새로운 열 만들기

경우에 따라 데이터에 있는 열을 이용하여 새로운 열을 만드는 경우가 있다. 예를 들어 radial 데이터에서 키와 몸무게를 이용하여 체질량지수(Body Mass Index, BMI)를 계산하려면 전처리 선택하기에서 **계산해서 새로운 열 만들기** 를 선택한 후 계산식에 BMI=weight/(height/100)^2 를 입력하고 **계산하기** 버튼을 누르면 된다.

제 11 절 서브그룹 만들기1

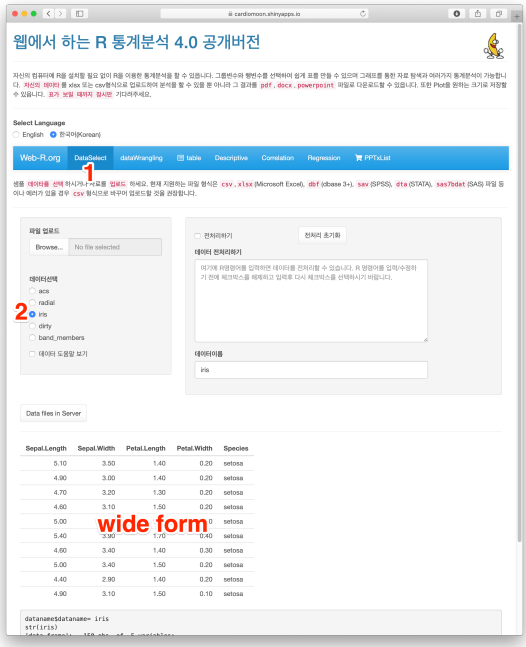
경우에 따라 연속형 변수를 기준으로 몇 개의 군으로 나누어 통계처리를 할 때가 있다. 예를 들어 radial 데이터에서 나이를 기준으로 세 개의 군으로 나누고자 하면 먼저 age를 선택하고 군갯수k에 3을 입력 후 **서브그룹 만들기** 버튼을 누르면 된다. 이 때에는 세 개의 군의 갯수가 최대한 비슷하도록 ageGroup이라는 변수에 1, 2, 3으로 입력된다.

제 12 절 서브그룹 만들기2

위의 방법은 우리가 정해준 군들의 갯수가 최대한 비슷하도록 기준을 임의로 정해 나누어주는 방법이다. 하지만 경우에 따라 기준이 정해져 있는 경우도 있다. 위에서 계산한 체질량지수(BMI)의 경우 저체중은 18.5 미만, 정상은 18.5-24.9, 과체중은 25-29.9, 비만은 30이상으로 그 기준이 정해져 있다. 이런 경우 cutpoint를 지정해 서브그룹을 나눌수 있는데 먼저 BMI를 선택하고(1) cutpoint에 0,18.5,25,30,100을 입력한 후(2) **서브그룹 만들기** 버튼을 누르면 네 개의 군으로 나누어진다. 이때 주의할 점은 cutpoint는 18.5,25,30의 세 개이지만 최소값, 최대값으로 대충의 값인 0과 100을 앞뒤로 넣어주어야 한다.

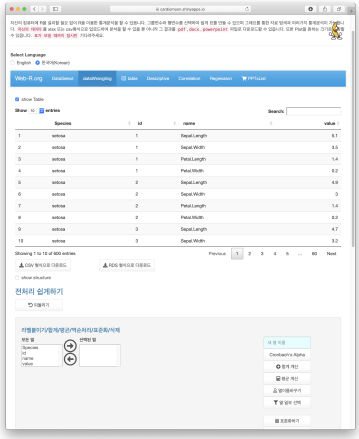
제 13 절 긴 형태로 바꾸기(pivot_longer)

통계처리나 그래프를 그리기 위해 데이터의 형태를 긴 형태(long form)로 바꾸어야 할 때가 있다. 먼저 dataSelect 메뉴에서 iris 데이터를 선택한다. 아래의 테이블에서 보이는 형태가 전형적인 넓은 형태(wide form)의 데이터로 이 중 Sepal.Length, Sepal.Width, Petal.Length 및 Petal.Width를 long from으로 바꾸어 본다.



dataWranging 메뉴에서 전처리 선택하기 중 **긴 형태로 바꾸기**를 선택하고(3) cols에 Sepal.Length, Sepal.Width, Petal.Length, Petal.Width를 선택하고(4) **pivot_longer** 버튼을 누른다(5).

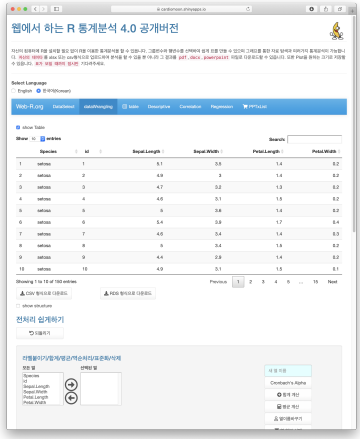
화면을 위로 올려 확인해보면 긴 형태(long form)의 데이터로 바뀐 것을 알 수 있다. 이를 다시 wide form 으로 바꾸려면 데이터의 id가 필요한데 웹R에서는 데이터의 id가 없는 경우 id를 추가해준다.



제 14 절 넓은 형태로 바꾸기(pivot_wider)

위 데이터를 다시 넓은 형태로 바꾸어 보자. 전처리 중 **넓은 형태로 바꾸기**를 선택하고 (1) pivot_wider 버튼을 누른다(2).

데이터의 형태가 다시 넓은 형태로 바뀌었다.



제 15 절 두 개의 데이터 병합하기

전처리에서 두 개의 데이터를 병합할 수 있다. 먼저 dataSelect 메뉴(1)에서 예제 데이터로 band_members를 선택한다(2).

dataWrangling 메뉴에서 화면을 아래쪽으로 이동한다. 두번째 데이터를 업로드하려면 **Browse...** 버튼을 이용하면 된다. 여기서는 Select Example 중 band_instruments(2) 를 선택한다. 데이터의 병합 방법은 left_join, right_join, inner_join, full_join, semi_join, anti_join, nest_join, bind_rows, bind_cols 등이 있는데 병합방법 선택(3)에서 선택할 수 있고 두 데이터의 구조 및 선택한 방법에 의한 병합의 결과를 미리 볼 수 있다(4). 미리 본 병합의 결과대로 병합을 하려면 **데이터 병합하기** 버튼을 누른다(5).

제 16 절 전처리가 끝난 데이터 다운로드

전처리가 끝나면 전처리가 끝난 데이터를 반드시 RDS형식으로 다운로드 받은 후 다시 업로드하여 통계분석을 진행하여야 한다. 그 이유는 전처리 하기 전의 데이터 이름과 전처리한 이후의 데이터 이름이 같기 때문에 이후 데이터 처리 과정에서 혼란이 빚어지기 때문이다.