한작지(Korean J. Crop Sci.), 64(4): 353~365(2019)
DOI : https://doi.org/10.7740/kjcs.2019.64.4.353

ISSN 0252-9777(Print)
ISSN 2287-8432(Online)

Original Research Article

밀 유전자원의 근적외선분광분석 예측모델에 의한 단백질 함량 변이분석

오세종[1,†] · 최유미[2] · 윤혜명[2] · 이수경[2] · 유은애[2] · 현도윤[2] · 신명재[2] · 이명철[1] · 채병수[3]

Statistical Analysis of Protein Content in Wheat Germplasm Based on Near-infrared Reflectance Spectroscopy

Sejong Oh[1,†], Yu Mi Choi[2], Hyemyeong Yoon[2], Sukyeung Lee[2], Eunae Yoo[2], Do Yoon Hyun[2], Myoung-Jae Shin[2], Myung Chul Lee[1], and Byungsoo Chae[3]

ABSTRACT A near-infrared reflectance spectroscopy (NIRS) prediction model was set to establish a rapid analysis system of wheat germplasm and provide statistical information on the characteristics of protein contents. The variability index value (VIV) of calibration resources was 0.80, the average protein content was 13.2%, and the content range was from 7.0% to 13.2%. After measuring the near-infrared spectra of calibration resources, the NIRS prediction model was developed through a regression analysis between protein content and spectra data, and then optimized by excluding outliers. The standard error of calibration, R^2, and the slope of the optimized model were 0.132, 0.997, and 1.000 respectively, and those of external validation results were 0.994, 0.191, and 1.013, respectively. Based on these results, a developed NIRS model could be applied to the rapid analysis of protein in wheat. The distribution of NIRS protein content of 6,794 resources were analyzed using a normal distribution analysis. The VIV was 0.79, the average protein was 12.1%, and the content range of resources accounting for 42.1% and 68% of the total accessions were 10-13% and 9.5-14.6%, respectively. The composition of total resources was classified into breeding line (3,128), landrace (2,705), and variety (961). The VIV in breeding line was 0.80, the protein average was 11.8%, and the contents of 68% of total resources ranged from 9.2% to 14.5%. The VIV in landrace was 0.76, the protein average was 12.1%, and the content range of resources of 68% of total accessions was 9.8-14.4%. The VIV in variety was 0.80, the protein average was 12.8%, and the accessions representing 68% of total resources ranged from 10.2% to 15.4%. These results should be helpful to the related experts of wheat breeding.

Keywords : near-infrared spectroscopy, normal distribution, protein content, wheat resources

밀은 쌀, 옥수수와 함께 세계 3대 주요 작물 중 하나이며 건조하고 척박한 환경에서도 잘 자라고 많은 노동력이 필요하지 않아 재배가 용이하다. 미네랄, 비타민 등 몸에 좋은 유용성분이 많아 식량작물로써의 조건을 가지고 있어 오래된 재배역사를 가지고 있다(Kang *et al*., 2016). 세계 인구의 약 30%가 주식으로 이용하고 있으며, 2017년 기준 국내의 1인당 밀가루 소비량은 32.4 kg으로 쌀 다음으로 소비량이 많다(Ministry of Agriculture, Food and Rural Affairs, 2018). 벼나 보리와는 달리 주로 밀가루 형태로 이용하는데, 이는 밀에 함유된 글루텐 단백질 복합체인 글리아딘과 글루테닌 단백질 분자들이 물을 흡수하여 점탄성을 갖는 반죽의 제조가 용이하기 때문에 다양한 식품의 원료로 이용되고 있다(Lim *et al*., 2007).

밀의 단백질 함량은 밀가루 반죽의 점성 및 탄성에 중요한 영향을 미치며 밀가루의 가공 용도를 구분하는 기준이 된다(Lee *et al*., 2002). 국외에서는 Heart bread and hard rolls

[1)]농촌진흥청 국립농업과학원 농업유전자원센터 연구관 (Senior Researcher, National Agrobiodiversity Center, National Institute of Agricultural Sciences, RDA, Jeonju 54874, Korea)
[2)]농촌진흥청 국립농업과학원 농업유전자원센터 연구사 (Researcher, National Agrobiodiversity Center, National Institute of Agricultural Sciences, RDA, Jeonju 54874, Korea)
[3)]농촌진흥청 국립농업과학원 농업유전자원센터 석사후전문연구원 (Master's Degree Researcher, National Agrobiodiversity Center, National Institute of Agricultural Sciences, RDA, Jeonju 54874, Korea)

†Corresponding author: Sejong Oh; (Phone) +82-63-238-4910; (E-mail) pleurotus@korea.kr
<Received 19 September, 2019; Revised 4 October, 2019; Accepted 22 October, 2019>

13.5 이상, Macaroni products 13% 이상, Pan bread 11.5-13.0%, Crackers 10-11%, Biscuits 9-11%, Cakes, Pies, cookies 8-10%로 규정하고 있다(Pomeranz, 1988). 국내는 박력분 강력분 중력분으로 구분해서 분류하고 있으며 박력밀가루는 반죽의 점성이 낮고 부드러워 제과용으로, 강력밀가루는 제빵용으로, 박력분과 강력분의 중간정도 글루텐 강도를 갖는 중력 밀가루는 제면용으로 사용된다(Kang *et al*., 2010).

밀 단백질 함량 분석에는 함유 질소량을 측정한 후 단백질 양으로 환산하는 마이크로 켈달분석법이 주로 사용되고 있으나, 이러한 방법은 많은 노력과 시간이 소요되는 단점이 있다(AACC, 2000). Near-infrared reflectance spectroscopy (NIRS)는 대량의 시료를 복잡한 전처리 과정 없이 다양한 성분을 빠른 시간 내 분석 가능한 장점이 있어 다양한 농산물 품질평가방법에 적용되고 있으며, 미국 공정법분석화학자협회(AOAC)와 미국 농무성(USDA)은 밀 단백질 함량 분석에 공식 이용하고 있다(Kim *et al.*, 2008a, 2008b; Song *et al*., 2006; Clarke *et al.*, 1992; Abrams *et al.*, 1987; Williams & Norris, 1987). 따라서 NIRS을 이용한 대량평가 분석방법은 농업의 다양한 분야에 적용되고 있는 바 제빵용 밀 품종의 질소 시비 방법과 글루텐 분포와의 관계(Cho *et al*., 2018), 국산 밀 품종의 농업형질과 내재해성과의 관계 분석(Shin *et al*., 2014), 온도 조건에 따라 벼 수량 및 수량 관련 요소와의 관계 분석(Lee *et al*., 2015), 다수성 벼 계통 선발을 위해 이삭 및 수량 관련 형질과의 관계 분석(Park *al*., 2015), 벼의 식미와 이화학적 성분들과의 관계 분석(Lee *et al*., 2012) 등의 연구가 보고되었다.

밀 단백질 습식분석 결과를 토대로 NIRS 프로그램에서 검량식을 작성하였으며 이들 좌표를 근거로하여 수동으로 회귀분석을 실시하여 통계적 분석적용을 이해하고 또한 신뢰성을 확인하고자 작성된 값들을 상호 비교하여 동일한 결과를 확인한 후 밀 단백질 성분 대량평가 기반이 되는 예측모델 자원집단을 확보하였다. 신속 정확한 대량평가 체계를 구축하여 농업유전자원센터에 보존되어있는 밀 유전자원성분 평가를 실시하여 그 결과를 단순한 도면으로 나타내고자 하였으며, 기초통계 기법을 이용하여 자원을 변이 별로 다양성 집단을 구분하고 이에 대한 밀 단백질 함량 분석정보를 관련 연구의 기초자료로 제공하고자 본 연구를 수행하였다.

재료 및 방법

검량식 자원 선정

밀 유전자원은 농업유전자원정보 통합관리시스템(GMS)을 이용하여 자원의 지리적 분포 특성이 잘 반영되도록 전체 자원을 아시아, 중동, 유럽, 아프리카, 아메리카, 오세아니아와 같이 6 그룹으로 분류한 후 각 그룹 간 구성 비율과

Table 1. Geographical origin distribution of wheat germplasm used for creating NIRS equation model.

Origin	Type	No. of accessions	Ratio of accessions (%)
China	Landrace	340	24.75
	Variety	105	
Korea	Landrace	108	10.29
	Variety	77	
Afghanistan	Landrace	180	10.01
United States of America	Landrace	1	7.90
	Variety	141	
India	Landrace	134	7.68
	Variety	4	
Turkey	Landrace	133	7.40
Spain	Landrace	48	2.73
	Variety	1	
Uzbekistan	Landrace	38	2.56
	Variety	8	
Ethiopia	Landrace	43	2.39
France	Landrace	5	2.06
	Variety	32	
Portugal	Landrace	37	2.06
Russia	Landrace	12	1.56
	Variety	16	
Bosnia and Herzegovina	Landrace	27	1.50
Tajikistan	Landrace	27	1.50
Italy	Landrace	3	1.39
	Variety	22	
Bulgaria	Variety	24	1.33
Pakistan	Landrace	19	1.11
	Variety	1	
Hungary	Landrace	2	1.06
	Variety	17	
The others	Landrace	39	10.72
	Variety	57	
	Unknown	97	
Total		1,798	100.00

*n = 1,798

각 그룹에 속하는 국가별 자원들의 구성 비율에 따라 2,098자원을 선발하였다. 선발자원 중 300자원을 제외한 1,798자원을 검량식 검정자원으로 선정하였다(Table 1). 각각의 종자 50립을 분쇄기(Bistro electric coffee grinder, Bodum®)를 사용하여 2,600 rpm 조건에서 약 2분간 분쇄하였다. 300 μm sieve를 사용하여 밀 겨가 제거된 균일한 밀가루 상태로 만든 후 근적외선영역의 스펙트럼 측정과 마이크로 켈달법에 의한 조단백질 함량 분석에 사용하였다. 종자 및 시료들은 13°C 저온저장고에 보관하여 수분함량을 14% 이하로 유지하였다. 수분함량은 105°C 온도조건에서 수분측정기(M0C63u, Shimadzu, Japan)를 사용하여 측정하였다.

조단백질 함량 분석

조단백질 함량은 micro Kjeldahl 질소정량법(AACC, 2000)에 따라 다음과 같이 측정하였다. 시료 0.5 g을 Kjeldahl 분해관에 넣고 진한 황산 12 ml를 가한 후 셀레늄분해촉매제(FOSS, Kjetabs Se/3,5, United Kingdom) 2알을 넣고 Kjeldahl 분해 및 포집 장치(FOSS, Tecator Digestor Auto & Scrubber, Suzhou, China)에 연결한 후 420°C에서 1시간동안 분해하였다. 분해 반응이 끝난 후 Kjeldahl 분해관을 실온에서 냉각하는 동안 잔여 분해가스를 방출시키고 Kjeldahl 자동 증류 및 분석 장치(FOSS, Kjeltec 8400, Hoganas, Sweden)에 연결하였다. 연결된 분해관에 40% NaOH 50 ml와 스팀을 가하여 분해 용액을 증류시켜 암모니아태 질소성분을 생성시킨 후 1% 붕산용액에 포집하였다. 붕산용액을 0.1 N HCl 표준용액으로 적정하여 질소함량(N%)을 구하고, 밀의 질소보정계수(5.83)를 곱하여 조단백질 함량(%)으로 환산하였다.

근적외선 스펙트럼 측정

밀 유전자원의 NIRS 측정은 근적외선 분광분석기(FOSS, XDS Rapid Content Analyzer, Hoganas, Sweden) 및 ISI scan (FOSS, ver. 4.2.0, 2007) 프로그램을 사용하였으며, sample cup에 시료 약 0.6 g을 채운후 sample cup backs로 시료내 공극을 최소화 시킨 후, 실온에서 동인 시료 당 2반복 scanning하여 근적외선 대역(700-2500 nm)의 스펙트럼을 얻고, 켈달분석법으로 얻어진 조단백질 함량을 입력한 후 검량식 작성에 활용하였다.

NIRS 검량식 작성 및 검증

습식분석으로 얻어진 단백질 함량이 저장된 근적외선 스펙트럼은 WINISI III project manager (FOSS, ver. 1.50e)의 standard normal variance와 detrend 기능을 이용하여 시료 입도 차이에서 발생하는 스펙트럼 산란현상의 보정, 수학적 처리, 회귀분석과 같은 일련의 처리과정을 거쳐 검량식 작성을 하였다. 수학적 처리는 1 ,4, 4, 1 (1st derivative, 4 nm gap, 4 point smooth, 1 point second smooth)을 적용하였다. 단백질 분석 값과 NIRS 측정값의 통계적 해석과 상관관계 분석을 통한 예측모델을 산출은 WINISI III project manager의 global equation 기능을 사용하였으며, 예측모델 중 단백질 습식 분석 값을 가장 잘 예측할 수 있는 모델은 R-square (RSQ, R^2), standard error of calibration (SEC), standard error of prediction (SEP), slope, bias, standard error of cross-validation (SECV) 및 one minus the ratio of unexplained variance to total variance (1-VR) 등의 통계지표를 고려하여 선별하였다. NIRS 검량식 작성 과정과 분석정보제공에 대한 체계를 Fig. 1에 나타냈다.

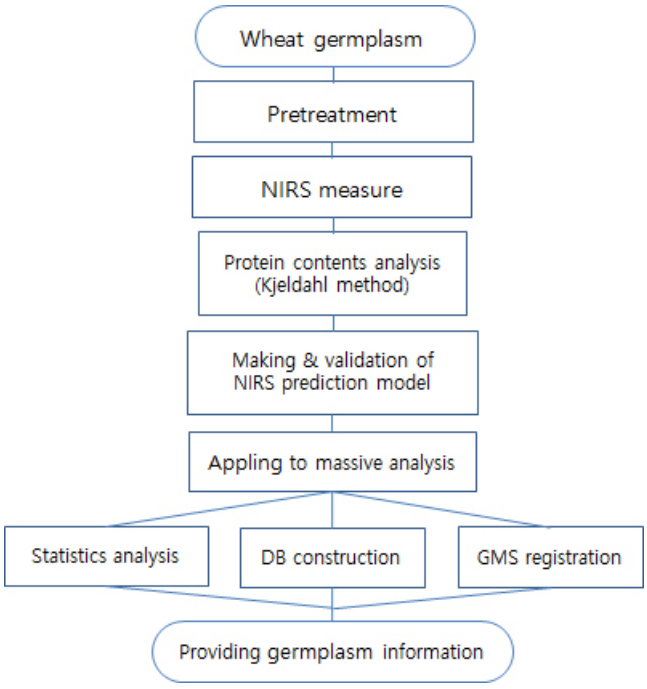


Fig. 1. Schematic design for system of rapid analysis evaluation.

검량식작성, 정규분포, 정규분포 표준화 및 다양성지수

검량식작성에 관련된 통계적이론과 분석정보를 바탕으로한 정규분포작성 그리고 자원정보 분석을 위한 정규분포 표준화에 대한 상세한 그림은 보충자료에 Supplement Figs. 1, 2, and 3 형식으로 예시자료로 추가하였다.

결과 및 고찰

NIRS 예측모델 설정

검량식 작성에 사용된 자원수와 R^2, slope 값의 통계지표를 고려하여 예측성능이 우수하다고 판단된 검량식을 선정하였고 이후 보완과정을 진행하였다. NIRS 분석방법은 습식분석에 비해 정확성은 낮기 때문에 검량식 자원들을 적정 농도 구간에서 균등한 자원 밀도를 나타내도록 구성하는 것이 중요하다(Kim *et al*., 2008a). 자원밀도가 낮은 구간은 자원을 추가 분석하여 검량식 작성에 사용하여 보완하였다. 이상치 자원을 검량식 작성과정에서 제거하여 주요 통계지표의 수치를 향상 시키는 방법으로 NIRS 예측모델을 최적화하였다. 검량식 작성과 보완에 사용된 1,798자원 중 최적화를 위해 111자원이 제외되었으며 최종적으로 1,687자원이 예측모델 작성에 사용되었다. 최적화된 NIRS 예측모델의 R^2 값은 0.997, SEC 값은 0.132, slope 값은 1.000이었다(Table 2). 검량식 작성에 사용된 밀 1,687자원들의 단백질 함량 분포 구간은 7.04-20.84%였고, 평균 함량은 13.2%, 표준편차는 2.6%, VIV 값은 0.80이었다. 전반적으로 검량식 자원구성이 정규분포를 이루어 예측모델 설정에 적합한 자원 집단이 구성되었다. 이는 대량평가체계를 구축하는 기반자원으로서의 역할을 충분히 할 수 있을 것으로 사료된다. 단백질 성분함량 13.0-15.5% 구간에서 타 함량 구간 대비 미약하게 낮은 확률밀도를 나타냈으나 해당 구간 내 자원수는 503자원으로 검량식 작성에 충분한 자원수로 판단되었다(Fig. 2).

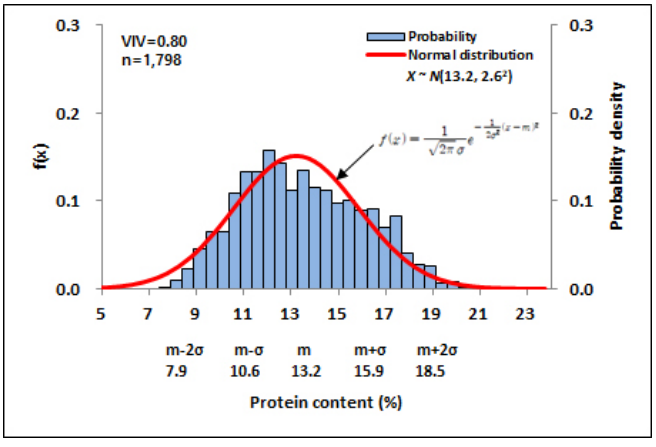


Fig. 2. Normal distribution of protein content in wheat germplasm for NIRS prediction model (n=1,798).

NIRS 예측모델 검정

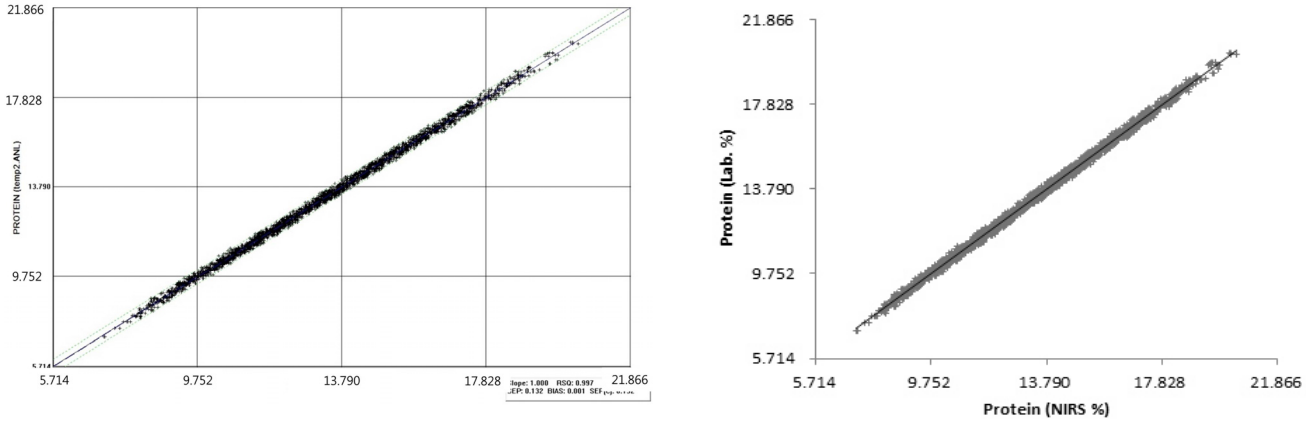
1-VR은 검량식 작성에 이미 이용된 자원을 재차 이용하여 정확도를 평가하는 역검정 방법이다. 간편하게 평가가 이뤄지는 장점이 있으나 1-VR 만으로는 개발된 NIRS 예측모델의 미지시료 분석 시 정확도를 평가하기에 부족하다(Bagchi *et al*., 2016). 따라서 검량식 작성에 사용된 자원 외의 별도자원들을 이용하여 NIRS 구동 프로그램의 external validation 기능을 이용하여 외부검증과정을 거쳤다. 검량식 작성에 사용된 1,798자원 이외의 300자원의 단백질 함량을 최적 예측모델이 적용된 NIRS를 이용하여 분석하였고, 이후 켈달법을 사용하여 조단백질 함량을 분석하였다. 검증자원 각각의 켈달분석 값을 NIRS 구동 프로그램의 lab data 항목에 입력하여 external validation set을 구성하였고, 구성된 set을 NIRS 예측모델에 적용하여 검정 결과 항목 각각의 통계지표들을 계산하였다. 검정결과와 예측모델의 통계지표들을 상호 비교해보면, r^2 (R^2) 값은 각각 0.994, 0.998이었고, SEP (SEC) 값은 각각 0.191, 0.132였고, slope 값은 각각 1.013, 1.000이었다. 각각의 통계지표들이 상호 유사한 수치를 나타내었고 r^2 (R^2), slope 값은 1에 가까웠고 SEP (SEC) 값은 비교적 낮은 수치를 나타냈다. 이는 밀 검량식 관련 선행 연구 결과와 동일하거나 그에 아주 근접한 통계치를 나타냈다(Shi *et al*., 2019; Zhang *et al*., 2017; Kim *et al*., 2016). 이상의 결과들을 종합해 볼 때 개발된 NIRS 예측모델에 의한 단백질 함량분석은 켈달분석과 높은 상관정도를 나타내며 분석정확도 또한 큰 차이가 없는 것으로 판단되었다(Table 2).

검량식 회귀분석

FOSS. ver. 1.50e에서 제공하는 WINISI III project manager 프로그램을 사용하여 작성된 NIRS 검량식 그래프와 EXCEL 프로그램을 사용하여 수동 방법으로 작성된 검량식 그래프를 상호 비교한 결과(Supplements 참조) 관련 통계지표들이 동일한 값을 나타내어 각기 다른 방법으로 작성된 검량식이 동일함을 확인하였다(Fig. 3).

Table 2. External validation result of NIRS optimized equation model for the analysis of protein content in wheat germplasm.

Constituent	External validation				NIRS optimized equation model			
	No.	R^2	SEP	Slope	No.	R^2	SEC	Slope
Protein	300	0.994	0.191	1.013	1,687	0.998	0.132	1.000



Protein content (%), N=1,687, R^2=0.998, Slope=1.000

(A) NIRS equation by using WINISI III program (B) NIRS equation by personal using EXCEL program

Fig. 3. Comparison of two methods of plotting NIRS equation graph between WINISI III program (A) and manual Excel program (B) based on the wheat germplasm.

Table 3. Classification of wheat flour based on protein content.

Commercial grade name	Wheat flour				
	Soft		Plain	Strong	
	Standard	High-grade	Standard	Standard	High-grade
Protein content (%)	8.5-9.0	below 8.5	9.0-10.5	over 11.0	10.5-11.0
Number of accessions	320	453	1,310	4,242	469
Ratio of accessions (%)	4.71	6.67	19.28	62.44	6.90

*n = 6,794

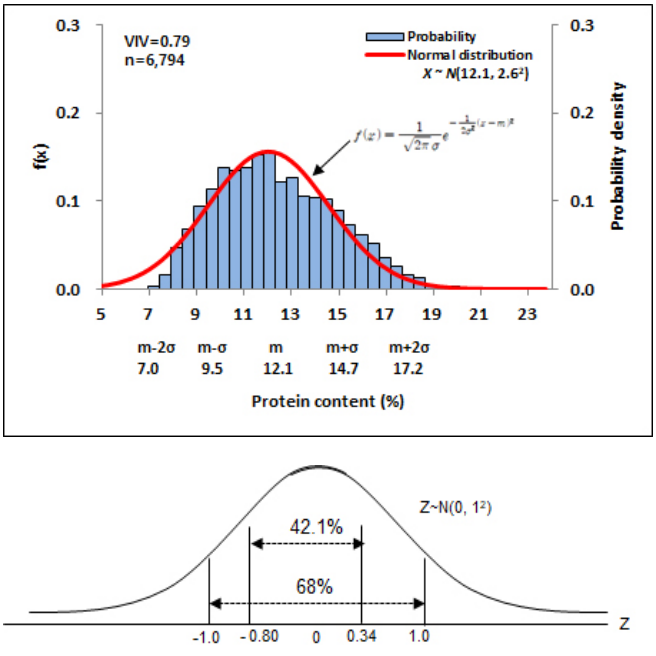


Fig. 4. Normal distribution and probability density of protein content in wheat germplasm.

밀 유전자원의 집단분석

시험에 사용된 밀 6,794자원의 단백질 함량분포를 정규분포곡선으로 나타냈다(Fig. 4). 단백질 12.5-13.5% 함량구간에서 정규분포함수에 비해 자원밀도가 다소 낮았다. 전반적으로 자원분포는 평균 12.1%, 표준편차 2.6%인 정규분포와 유사하였고, 자원의 다양성 지수는 0.79였다. 전체 자원수의 약 68%에 해당되는 자원들이 9.5-14.6% 함량구간에 속했다(Fig. 4). 밀가루 등급 기준으로 밀 6,794자원을 구분하면 전체 자원의 62.44%인 4,242자원이 강력분 표준등급이었으며 전체 자원의 19.28%인 1,310자원이 중력분 표준 등급이었다(Table 3).

밀 유전자원의 type별 특성

유전자원센터에 보존 중인 전체 밀 유전자원은 약 20,000여 자원이 되지만(2018년 기준) 이 중에서 원산지 미상자원을 제외하면 총 6,794자원이 되며 이는 밀 유전자원의 약 34%에 해당한다. 평균은 12.1이며, 표준편차는 2.6이고, ±σ까지 차지하는 비율은 68%로써 4,619자원이 분포되어있으

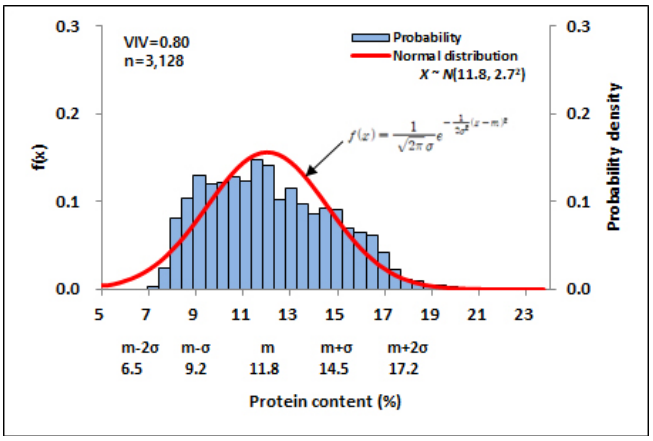
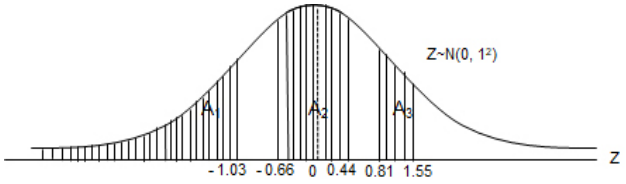


Fig. 5. Normal distribution of wheat protein content in breeding line obtained from NIRS analysis.



Contents	Proportion (%)	No. of accession
X≤9.0	15.1(A_1)	472
10.0≤X≤13.0	41.5(A_2)	1,298
14.0≤X≤16.0	14.8(A_3)	463
Total accessioons : 3,128		

Fig. 6. Calculation of accessions for wheat protein content in breeding line using standard normal distribution curve.

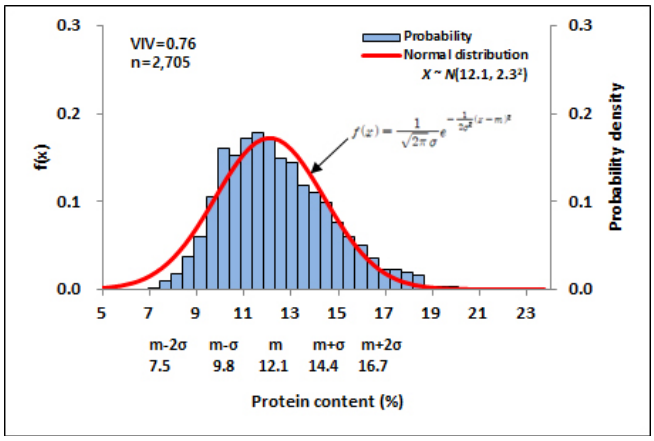
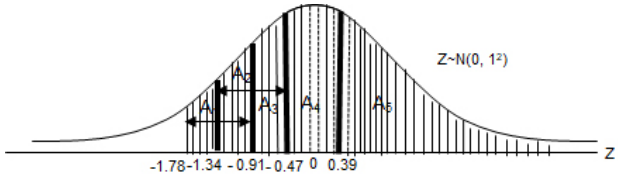


Fig. 7. Normal distribution of wheat protein content in landrace obtained from NIRS analysis.



Contents	Proportion (%)	No. of accession
8.0≤X≤10.0	14.3(A_1)	386
9.0≤X≤11.0	22.9(A_2)	619
10.0≤X≤11.0	13.7(A_3)	370
11.5≤X≤13.0	33.2(A_4)	898
13.0≤X	34.8(A_5)	941
Total accessioons : 2,705		

Fig. 8. Calculation of accessions for wheat protein content in landrace using standard normal distribution curve.

며, 다양성지수는 0.79인 다소 높은 다양성을 보였다(Fig. 4). 전체자원을 세부적으로 Type별로 구분해 보면 육성계통 3,128자원, 재래종 2,705자원, 육성품종 961자원으로 나눌 수 있으며 이들을 대상으로 단백질 함량 분포 특성을 비교하였다(Fig. 5, Fig. 7, Fig. 9).

육성계통 자원은 단백질 7-7.5%, 12.5-14.0% 함량구간에서 정규분포함수에 비해 자원밀도가 낮았고, 8.0-9.5% 함량구간에서 정규분포함수에 비해 자원밀도는 높았다. 전반적으로 자원분포는 평균 11.8%, 표준편차 2.7%인 정규분포와 유사하였고, 자원의 다양성 지수는 0.80이었다(Fig. 5). 표준정규분포 곡선을 이용한 자원분포 분석에서 박력분에 해당하는 9.0% 이하 범위는 전체자원에서 15.1%인 472자원을 차지하고 있으며, 10≤x≤13% 사이에는 41.5%인 1,298자원 그리고 14≤x≤16% 사이에는 14.8%인 463자원을 차지했다(Fig. 6). 이와 같이 표준정규 분포곡선을 이용하면 어느 구간이든지 단백질 함량 자원을 손쉽게 구할 수 있으며 또한 보존자원수를 산출 해 낼 수 있다. 이는 앞으로 프로그램화해서 실시간으로 수요자에게 정보를 제공하게 되리라 사료된다.

재래종 자원은 7.0-8.5%, 12.5-13.0%, 13.5-14.0% 함량구간에서 정규분포함수에 비해 자원밀도가 낮았고 11.0-12.0% 함량구간에서 정규분포함수에 비해 자원밀도는 높았다. 전반적으로 자원분포는 평균 12.1%, 표준편차 2.3%인 정규분포와 유사하였고, 자원의 다양성 지수는 0.76이었다(Fig. 7). Pomeranz의 구분기준에 근거로한 재래종 밀 유전자원 분포분석은 13.0% 이상은 전체자원의 34.8%인 941자원, 11.5-13.0% 범위는 전체자원의 33.2%인 898자원, 10-11% 범위는 전체자원의 13.7%인 370자원, 9-11% 범위는 전체

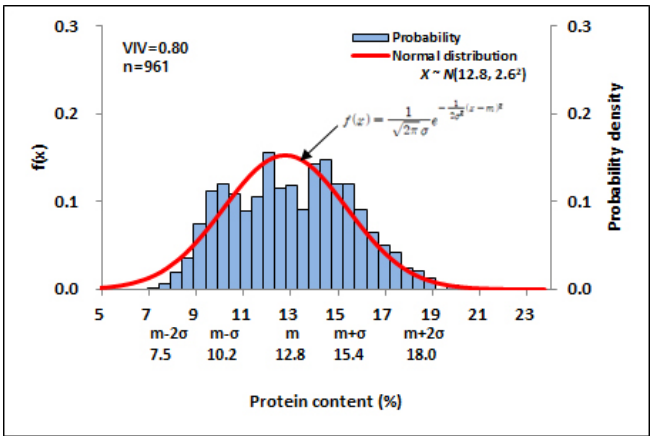
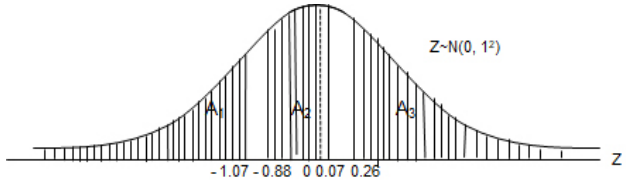


Fig. 9. Normal distribution of wheat protein content in variety obtained from NIRS analysis.



Contents	Proportion (%)	No. of accession
X≤10.0	14.2(A_1)	136
10.5≤X≤13.0	33.8(A_2)	324
13.5≤X	39.7(A_3)	381
Total accessioons : 961		

Fig. 10. Calculation of accessions for wheat protein content in variety using standard normal distribution curve.

자원의 22.9%인 619자원, 8-10% 범위는 전체자원의 14.3%인 386자원을 차지하는 것으로 나타났다(Fig. 8). 재래종 자원은 육종소재로 사용될 가능성이 많으므로 자원분석을 용도별로 빵용, 국수용, 제과용 등의 구분에 따라 여러 구간으로 나누어 봄으로써 관련 연구자에게 유용한 정보가 제공될 것으로 사료된다.

육성품종 자원은 7.0-8.0%, 11.0-12.0%, 12.5-14.0% 함량구간에서 정규분포함수에 비해 자원밀도가 구간별로 조금씩 차이가 있었으며 9.0-10.5%, 14.5-16.5% 함량구간에서 정규분포함수에 비해 자원밀도는 높았다. 전반적으로 자원분포는 평균 12.8%, 표준편차 2.6%인 정규분포와 유사하였고, 자원의 다양성 지수는 0.80이었다(Fig. 9). 밀 육성품종은 x≤10.0 범위는 전체자원의 14.2%인 136자원, 10.5≤x≤13.0 범위는 전체자원의 33.8%인 324자원, 13.5≤x 범위는 39.7%인 381자원을 각각 차지하였다(Fig. 10). 밀 육성품종은 13.5%이상의 자원이 다소 많은 비중을 차지하고 있는데 이는 대부분이 식량자원으로써 빵용으로 사용하는 밀이 육성되었다고 볼 수 있다. 제시된 그림(Fig. 9)에서와 같이 육성품종은 11-14% 사이에 자원이 정규분포와 일치하지 않는 것은 이상적인 자원집단분포와 약간의 차이를 보인다. 이것은 이론과 실제가 당연히 차이가 나타나게 마련이며 실질적으로 어떤 집단을 분석해도 이론적인 분포상을 나타내는 결과는 극히 드물며 이를 위해서 약간의 교정이 있을 수는 있지만 실제 집단분석에서는 그와 같은 그림은 볼 수 없는 현상이 대부분이라고 생각된다. 함수식이 제공하는 정규분포와 수집자원 히스토그램의 모양이 일치한다면 이상정인 수집자원 집단을 형성하고 있다고 해석할 수 있다. 히스토그램이 정규분포곡선과 차이를 보이는 부분은 앞으로 수집자원의 보정이 필요하다는 것을 제시할 수도 있으며 이를 최소화함으로써 프로그램화하는데 기초자료로 제공될 가능성이 있을 것으로 사료된다. 유전자원이 자원으로써 중요성을 가지기 위해서는 자원의 다양성을 얼마나 갖추고 있느냐에 달려 있다고 생각된다. 따라서 다양성지수가 의미 있는 수치로 생각되며 그에 따른 표준편차 또한 동일한 효과를 가지는 자료라고 사료된다. 따라서 그에 관한 정보를 정규분포그래프 에 모두 표시하였다.

적 요

본 연구는 근적외선 분광분석기(NIRS) 예측모델을 설정하여 유전자원 대량분석 체계를 확립하고 그에 따른 국내외 밀 자원의 단백질 함량에 관한 기초 정보를 제공하고자 하였다.

1. 농업유전자원센터에 보유하고 있는 20,000여 자원 중 1,798자원을 검량 자원으로 선발하였다. 검량자원의 NIR 스펙트럼을 측정하였고, 단백질 함량 습식분석 데이터 입력 등 일련의 통계적 처리 과정을 거쳐 NIRS 예측모델을 설정했다. 검량 자원의 다양성 지수는 0.80이었고, 습식 분석법에 의한 단백질 평균은 13.2%, 함량 구간은 7.0-20.8%였다. 최적화된 NIRS 모델의 R^2, SEC, Slope은 0.997, 0.132, 1.000이었다. 300자원을 사용하여 외부검정 과정을 실시하였고 R^2, SEP, Slope은 0.994, 0.191, 1.013이었다. 최적화된 NIRS 모델과 외부검정 결과의 통계치가 상호 유사하였고, 1에 가까운 R^2와 Slope 값, 낮은 SEC와 SEP 값을 볼 때 본 연구에서 설정한 NIRS 모델은 습식 분석법을 대체하여 밀 자원의 단백질 함량 분석에 적용 가능할 것으로 판단되었다.

2. 국내외 수집된 밀 6,794자원의 NIRS 단백질 함량 측정값을 정규분포로 작성하여 특성을 파악했다. 자원의 다양성 지수는 0.79, 단백질 평균은 12.1%, 전체 자원의 임의구간 42.1% 단백질 함량자원 범위는 10-13%이었으며, 68.0%를 차지하는 자원들의 단백질 함량 범위는 9.5-14.7%였다.
3. 전체 6,794자원의 품종 집단 구성은 육성계통 3,128자원, 재래종 2,705자원, 육성품종 961자원이었다. 육성계통 자원의 다양성 지수는 0.80, 단백질 평균은 11.8%, 전체 자원의 68%를 차지하는 자원들의 함량 범위는 9.2-14.5%였다. 재래종 자원의 다양성 지수는 0.76, 단백질 평균은 12.1%, 전체 자원의 68.0%를 차지하는 자원들의 함량 범위는 9.8-14.4%였다. 육성품종 자원의 다양성 지수는 0.80, 단백질 평균은 12.8%, 전체 자원의 68.0%를 차지하는 자원들의 함량 범위는 10.2-15.4%였다. 재래종 자원은 가장 낮은 다양성 지수를 나타냈고, 육성계통과 육성품종은 동일한 다양성 지수를 나타냈다. 육성계통은 가장 낮은 단백질 평균을 나타냈고, 육성품종은 가장 높은 단백질 평균을 나타냈다.

사 사

본 연구는 농촌진흥청 농업과학기술연구 개발사업(과제번호: PJ01353904)의 지원에 의해 이루어졌습니다.

인용문헌(REFERENCES)

AACC. 2000. Approved methods of the American Association of Cereal Chemists In : St. Paul, MN, 10th edn. USA.

Abrams, S. M., J. S. Shenk, M. O. Westerhaus, and F. E. Barton. 1987. Determination of forage quality by near-infrared reflectance spectroscopy: Efficiency of broad based calibration equations. J. Diary Sci. 70 : 806-813.

Bagchi, T. B., S. Sharma, and K. Chattopadhyay. 2016. Development of NIRS models to predict protein and amylose content of brown rice and proximate compositions of rice bran. Food Chemistry 191 : 21-27.

Clarke, M. A., E. R. Arias, and C. McDonald-Lewis. 1992. Near Infrared Analysis in The Sugarcane Factory, Ruspam Commun. Inc., Sugary Azucar Press, LA (USA). pp. 244-264.

Cho, S. W., T. G. Kang, C. S. Park, J. H. Son, C. H. Choi, Y. K. Cheong, Y. M. Yoon, K. H. Kim, and C. S. Kang. 2018. Influence of different nitrogen fertilizer application levels and application timing on gluten fraction and bread loaf volume during grain filing. Korean J. Crop Sci. 63(3) : 229-238.

Kang, C. S., C. S. Park, J. C. Park, H. S. Kim, Y. K. Cheong, K. H. Kim, K. J. Kim, K. H. Park, and J. G. Kim. 2010. Flour characteristics and end-use quality of Korean wheat cultivars II. End-use properties. Korean J. Breed. Sci. 42(1) : 75-86.

Kang, C. S., Y. K. Cheong, and B. K. Kim. 2016. Current situation and prospect of Korean wheat industry. Food Industry and Nutrition 21(2) : 20-24.

Kim, J. S., M. H. Song, J. E. Choi, H. B. Lee, and S. N. Ahn. 2008a. Quantification of protein and amylose contents by near-infrared reflectance spectroscopy in aroma rice. Korean J. Food Sci. Technol. 40(6) : 603-610.

Kim, J. S., Y. H. Cho, J. G. Gwag, K. H. Ma, Y. M. Choi, J. B. Kim, J. H. Lee, T. S. Kim, J. K. Cho, and S. Y. Lee. 2008b. Quantitative analysis of amylose and protein content of rice germplasm in RDA-genebank by near-infrared reflectance spectroscopy. Korean J. Crop Sci. 53(2) : 217-223.

Kim, K. H., C. S. Kang, I. D. Choi, H. S. Kim, J. N. Hyun, and C. S. Park. 2016. Analysis of grain characteristics in Korean wheat and screening wheat for quality using near infrared reflectance spectroscopy. Korean J. Breed. Sci. 48(4) : 442-449.

Lee, C. K., J. H. Nam, M. S. Kang, B. C. Ku, J. C. Kim, K. G. Park, M. W. Park, and Y. H. Kim. 2002. Current wheat quality criteria and inspection system of major wheat producing countries. Korean Crop J. Sci. 47 : 63-94.

Lee, J. Y., H. J. Lee, J. H. Cho, S. Y. Kim, C. S. Kim, Y. B. Sohn, U. S. Yeo, C. W. Lee, and M. H. Nam. 2012. Analysis of eating quality in recombinant inbred lines and selection of elite line with glutelin content in rice. Korean J. Breed. Sci. 44(2) : 136-141.

Lee, K. J., D. J. Kim, H. Y. Ban, and B. W. Lee. 2015. Genotypic differences in yield and yield-related elements of rice under elevated air temperature conditions. Korean Journal of Agricultural and Forest Meteorology 17(4) : 306-316.

Lim, E. Y., H. K. Chang, and Y. S. Park. 2007. Physicochemical properties and product potentiality of soft wheats. Korean J. Food Sci. Technol. 39(4) : 412-418.

Ministry of Agriculture, Food and Rural Affairs. 2018. Agriculture, food and rural affairs statistics yearbook. MAFRA. Sejong. Korea. 214-215.

Pomeranz, Y. 1988. Criteria of wheat quality. Pages 15-45 in: Wheat Chemistry and Technology. American Association of Cereal Chemists. ST. Paul. Mn.

Park, H. S., K. Y. Ha, K. Y. Kim, W. J. Kim, J. K. Nam, M. K. Baek, J. J. Kim, J. M. Jeong, Y. C. Cho, J. H. Lee, B. K. Kim, and S. N. Ahn. 2015. Development of high-yielding rice lines and analysis of panicle and yield-related traits using doubled haploid lines derived from the cross between deuraechan and boramchan, high-yielding *japonica* rice cultivars in Korea. Korean J. Breed. Sci. 47(4) : 384-402.

Shin, S. H., K. H. Kim, J. H. Son, C. S. Kang, Y. K. Cheong, C. K. Le, J. C. Park, and C. S. Park. 2014. Analysis of semi-dwarf gene (Rht) construction and its relationship with agronomic

characteristics, pre-harvest spouting, and fusarium head blight in Korean wheat cultivar. Journal of Agriculture & Life Sciences 45(1) : 72-79.

Shi, H., Y. Lei, L. L. Prates, and P. Yu. 2019. Evaluation of near-infrared (NIR) and fourier transform mid-infrared (ATR-FT/MIR) spectroscopy techniques combined with chemometrics for the determination of crude protein and intestinal protein digestibility of wheat. Food Chemistry 272 : 507-513.

Williams, P. and K. Norris. 1987. Near-Infrared Technology in Agricultural and Food Industries. American Association of Cereal Chemists, Inc., MN (USA). p. 330.

Zhang, Y., L. Luo, J. Li, S. Li, W. Qu, H. Ma, A. O. Oladejo, and X. Ye. 2017. In-situ and real-time monitoring of enzyme process of wheat gluten by miniature fiber NIR spectrometer. Food Research International 99 : 147-154.

SUPPLEMENT

검량식작성 회귀분석

$S_i = \{y_i - (a+bx_i)\}^2$

$S = \sum_{i=1}^{n} \{y_i - (a+bx_i)\}^2$

S를 a에 대해서 편미분하고 $\frac{\partial S}{\partial a} = -2\sum_{i=1}^{n}(yi - a - bxi) = 0$

b에 대해서 편미분하면 $\frac{\partial S}{\partial b} = -2\sum_{i=1}^{n} x_i(y_i - a - bx_i) = 0$

$\sum_{i=1}^{n} y_i - Na - \sum_{i=1}^{n} x_i b = 0$ 식 (1)

$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i a - \sum_{i=1}^{n} x_i^2 b = 0$ 식 (2)

식 (1) $\Rightarrow$ $Na + \sum_{i=1}^{n} x_i b = \sum_{i=1}^{n} y_i$

식 (2) $\Rightarrow$ $\sum_{i=1}^{n} x_i a + \sum_{i=1}^{n} x_i^2 b = \sum_{i=1}^{n} x_i y_i$

위의 식 (1)과 식 (2)의 연립방정식 $\Rightarrow$ 행렬로

$$\begin{bmatrix} N & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix}$$

a, b를 구하기 위하여 Cramer's rule 적용

*Determinant $D = N\Sigma x_i^2 - (\Sigma x_i)^2$, Da, Db $\Rightarrow$ a = $\frac{Da}{D}$, b = $\frac{Db}{D}$

Da = $\begin{bmatrix} \Sigma y_i & \Sigma x_i \\ \Sigma x_i y_i & \Sigma x_i^2 \end{bmatrix} = \Sigma y_i \Sigma x_i^2 - \Sigma x_i y_i \Sigma x_i$ Db = $\begin{bmatrix} N & \Sigma y_i \\ \Sigma x_i & \Sigma x_i y_i \end{bmatrix} = N\Sigma x_i y_i - \Sigma x_i \Sigma y_i$

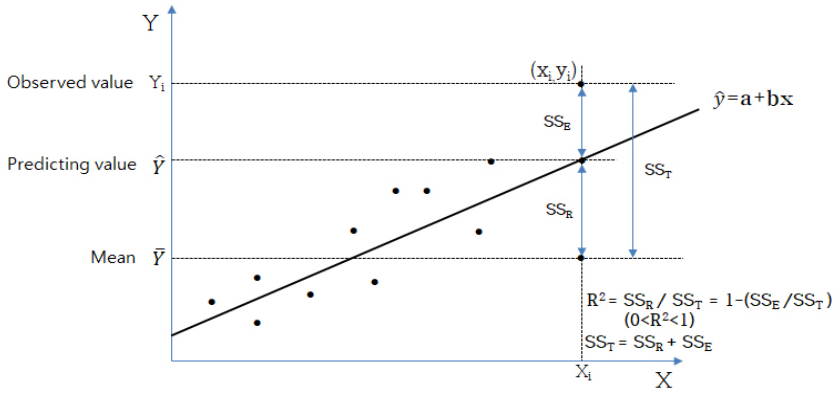
$$a = \frac{\Sigma x_i^2 \Sigma y_i - \Sigma x_i \Sigma x_i y_i}{N\Sigma x_i^2 - (\Sigma x_i)^2}, \quad b = \frac{N\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{N\Sigma x_i^2 - (\Sigma x_i)^2}$$

$$\therefore \hat{Y} = a + bX = \frac{\Sigma x_i^2 \Sigma y_i - \Sigma x_i \Sigma x_i y_i}{N\Sigma x_i^2 - (\Sigma x_i)^2} + \frac{N\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{N\Sigma x_i^2 - (\Sigma x_i)^2} X$$

$$R^2 = \frac{SS_R}{SS_T} = \frac{\sum(\hat{Y}_i - \overline{Y})^2}{\sum(Y_i - \overline{Y})^2} = \frac{\sum(\hat{Y}_i^2 - 2\hat{Y}_i\overline{Y} + \overline{Y}^2)}{\sum(Y_i^2 - 2Y_i\overline{Y} + \overline{Y}^2)} = \frac{\sum \hat{Y}_i^{\,2} - 2\sum \hat{Y}_i \frac{\sum \hat{Y}_i}{N} + \sum \frac{(\sum \hat{Y}_i)^2}{N^2}}{\sum Y_i^2 - 2\sum Y_i \frac{\sum Y_i}{N} + \sum \frac{(\sum Y_i)^2}{N^2}}$$

$$= \frac{\sum \hat{Y}_i^{\,2} - 2\sum \hat{Y}_i \frac{\sum \hat{Y}_i}{N} + \frac{(\sum Y)^2}{N}}{\sum Y_i^2 - 2\sum Y_i \frac{\sum Y_i}{N} + \frac{(\sum Y_i)^2}{N}} = \frac{\sum \hat{y_i}^2 - 2\frac{(\sum \hat{Y}_i)^2}{N} + \frac{(\sum \hat{Y}_i)^2}{N}}{\sum Y_i^2 - 2\frac{(\sum Y_i)^2}{N} + \frac{(\sum Y_i)^2}{N}} = \frac{\sum \hat{Y}_i^{\,2} - \frac{(\sum \hat{Y}_i)^2}{N}}{\sum Y_i^2 - \frac{(\sum Y_i)^2}{N}} = \frac{N\sum \hat{Y}_i^{\,2} - (\sum \hat{Y}_i)^2}{N\sum Y_i^2 - (\sum Y_i)^2}$$

R^2, SS_T, SS_R, SS_E들 간의 관계에 관한 모식도는 Supplement Fig. 1과 같다.



Supplement Fig. 1. Schematic diagram on two dimension for analysis of regression.

$$CP_{xy} = \sum(X_i-\overline{X})(Y_i-\overline{Y}) = \sum(X_iY_i - X_i\overline{Y} - Y_i\overline{X} + \overline{X}\overline{Y}) = \sum X_iY_i - \overline{Y}\sum X_i - \overline{X}\sum Y_i + \sum \overline{X}\overline{Y} \ (\because \overline{X}, \overline{Y}\text{=constant})$$

$$= \sum X_iY_i - \frac{\sum Y_i}{N}\sum Y_i - \frac{\sum X_i}{N}\sum Y_i + \sum \frac{\sum X_i \sum Y_i}{N^2} = \sum X_iY_i - 2\frac{\sum X_i \sum Y_i}{N} + \frac{\sum X_i \sum Y_i}{N}$$

$$= \sum X_iY_i - \frac{\sum X_i \sum Y_i}{N}$$

$$Cov(x,y) = \frac{CP_{xy}}{N-1} = \frac{\sum(X_i-\overline{X})(Y_i-\overline{Y})}{N-1} = \frac{\sum X_iY_i - \frac{\sum X_i \sum Y_i}{N}}{N-1}$$

$$\left(\sqrt{s_x^2}\sqrt{s_y^2}\right)^2 = (s_x^2 s_y^2)^2 = \sum(X_i-\overline{X})^2 \sum(Y_i-\overline{Y})^2 = \left\{\sum X_i^2 - \frac{(\sum X_i)^2}{N}\right\}\left\{\sum Y_i^2 - \frac{(\sum Y_i)^2}{N}\right\}$$

$$\Rightarrow \sum(X_i-\overline{X})^2 = \sum(X_i^2 - 2X_i\overline{X} + \overline{X}^2) = \sum X_i^2 - 2N\overline{X}\frac{\sum X_i}{N} + \sum \overline{X}^2 = \sum X_i^2 - 2N\overline{X}^2 + N\overline{X}^2$$

$$= \sum X_i^2 - N\overline{X}^2 = \sum X_i^2 - N\left(\frac{\sum X_i}{N}\right)^2 = \sum X_i^2 - \frac{(\sum X_i)^2}{N}$$

$$\sum(Y_i-\overline{Y})^2 = \sum(Y_i^2 - 2Y_i\overline{Y} + \overline{Y}^2) = \sum Y_i^2 - 2N\overline{Y}\frac{\sum Y_i}{N} + \sum \overline{Y}^2 = \sum Y_i^2 - 2N\overline{Y}^2 + N\overline{Y}^2$$

$$= \sum Y_i^2 - N\overline{Y}^2 = \sum Y_i^2 - N\left(\frac{\sum Y_i}{N}\right)^2 = \sum Y_i^2 - \frac{(\sum Y_i)^2}{N}$$

따라서, R^2의 정의식과 계산식은 다음과 같다.

$$R^2 = r^2 = \left\{\frac{Cov(x,y)}{\sqrt{s_x^2}\sqrt{s_y^2}}\right\}^2 = \left\{\frac{\frac{\sum(X_i-\overline{X})(Y_i-\overline{Y})}{N-1}}{\sqrt{\frac{\sum(X_i-\overline{X})^2}{N-1}}\sqrt{\frac{\sum(Y_i-\overline{Y})^2}{N-1}}}\right\}^2$$

$$= \frac{\{\sum(X_i-\overline{X})(Y_i-\overline{Y})\}^2}{\sum(X_i-\overline{X})^2\sum(Y_i-\overline{Y})^2} = \frac{\left(\sum X_iY_i - \frac{\sum X_i\sum Y_i}{N}\right)^2}{\left\{\sum X_i^2 - \frac{(\sum X_i)^2}{N}\right\}\left\{\sum Y_i^2 - \frac{(\sum Y_i)^2}{N}\right\}} = \frac{(N\sum X_iY_i - \sum X_i\sum Y_i)^2}{\{N\sum X_i^2 - (\sum X_i)^2\}\{N\sum Y_i^2 - (\sum Y_i)^2\}}$$

Supplement Table 1. Treatment of statistics data for manual calculation.

Method	Regression analysis					Correlation coefficient	
Definition of R^2	$\frac{SS_R}{SS_T}=\frac{\sum(\hat{Y}_i-\overline{Y})^2}{\sum(Y_i-\overline{Y})^2}=\frac{N\sum\hat{Y}_i{}^2-(\sum\hat{Y}_i)^2}{N\sum Y_i^2-(\sum Y\hat{i})^2}$					$\left\{\frac{Cov(x_i,y_i)}{\sqrt{S_x^2\times S_y^2}}\right\}^2=\frac{\left(N\sum X_iY_i-\sum X_i\sum Y_i\right)^2}{\left\{N\sum X_i^2-(\sum X_i)^2\right\}\left\{N\sum Y_i^2-(\sum Y_i)^2\right\}}$	
Number of observed values (N)	X_i	Y_i	X_i^2	Y_i^2	X_iY_i	$\hat{Y}_i$	$\hat{Y}_i^2$
1	7.14	7.36	7.14^2	7.36^2	7.14*7.36	7.26	7.26^2
2	8.20	8.22	8.20^2	8.22^2	8.20*8.22	8.32	8.32^2
3	9.01	9.32	9.00^2	9.32^2	9.00*9.32	9.12	9.12^2
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
50	17.82	18.39	17.82^2	18.39^2	17.82*18.39	17.91	17.91^2
$\sum X_i$	624.569						
$\sum Y_i$		629.684					
$\sum X_i^2$			8276.82232				
$\sum Y_i^2$				8411.21624			
$\sum X_iY_i$					8339.25103		
N		50		50			
$\overline{Y}=\frac{\sum Y_i}{N}$		12.59					
$\sum\hat{Y}_i$						629.68583	
$\sum\hat{Y}_i^2$							8402.25636

$\hat{Y}$

b (slope) $=\frac{N\sum X_iY_i-\sum X_i\sum Y_i}{N\sum X_i^2-(\sum X_i)^2}$

$=\frac{50\times8339.25-624.57\times629.68}{50\times8276.82-(624.57)^2}$

$=0.99692$

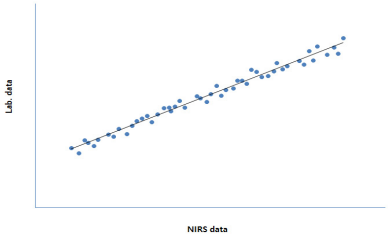
a (intercept)

$=\frac{\sum X_i^2\sum Y_i-\sum X_i\sum X_iY_i}{N\sum X_i^2-(\sum X_i)^2}$

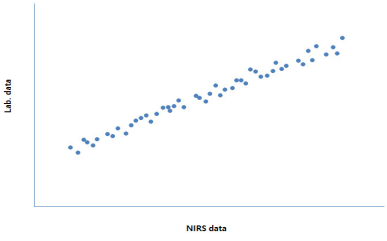
$=\frac{8276.82\times629.68-624.57\times8339.25}{50\times8276.82-(624.57)^2}$

$=0.14081$

$\hat{Y}=0.14081+0.99692x$



<Scatter plot with regression equation>



<Scatter plot>

R^2

$\frac{SS_R}{SS_T}=\frac{\sum(\hat{Y}_i-\overline{Y})^2}{\sum(Y_i-\overline{Y})^2}=\frac{\sum\hat{Y}_i{}^2-\frac{\left(\sum\hat{Y}_{i_i}\right)^2}{N}}{\sum Y_i^2-\frac{\left(\sum Y_i\right)^2}{N}}=\frac{8402.25636-\frac{(629.68583)^2}{50}}{8411.21624-\frac{(629.684)^2}{50}}$

$=0.98128$

$\left\{\frac{Cov(x_i,y_i)}{\sqrt{S_x^2\times S_y^2}}\right\}^2=\frac{\left\{\sum(X_i-\overline{X})(Y_i-\overline{Y})\right\}^2}{\sum(X_i-\overline{X})^2\sum(Y_i-\overline{Y})^2}$

$=\frac{\left(\sum X_iY_i-\frac{\sum X_i\sum Y_i}{N}\right)^2}{\left\{\sum X_i^2-\frac{(\sum X_i)^2}{N}\right\}\left\{\sum Y_i^2-\frac{(\sum Y_i)^2}{N}\right\}}$

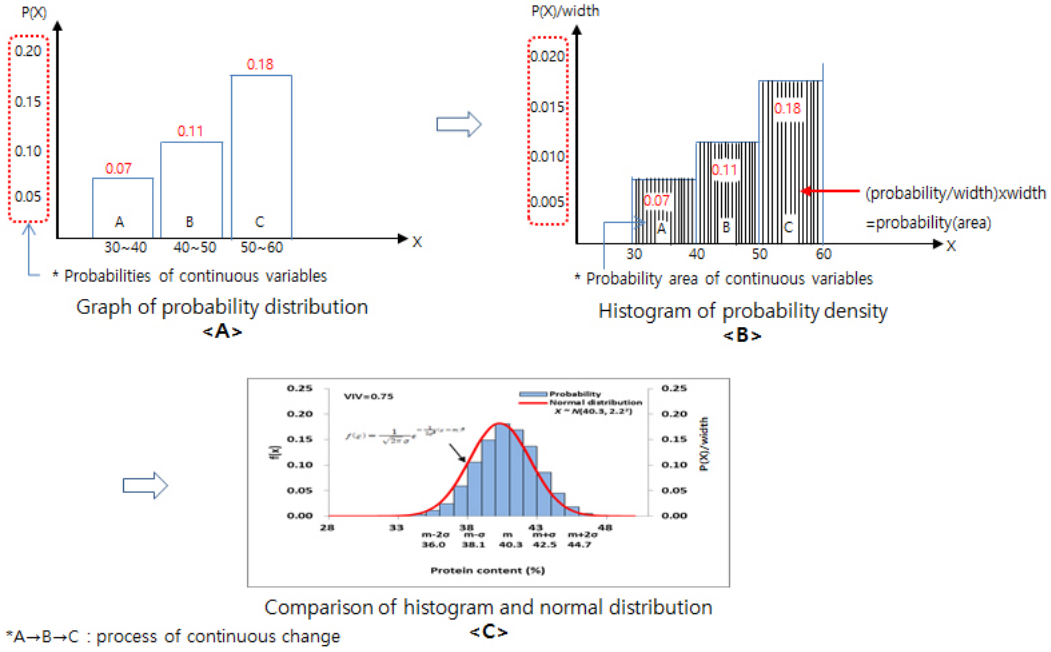
$=\frac{\left(8339.25103-\frac{624.569\times629.684}{50}\right)^2}{\left\{8276.82232-\frac{(624.569)^2}{50}\right\}\left\{8411.21624-\frac{(629.684)^2}{50}\right\}}$

$=0.98128$

다양성 지수 계산

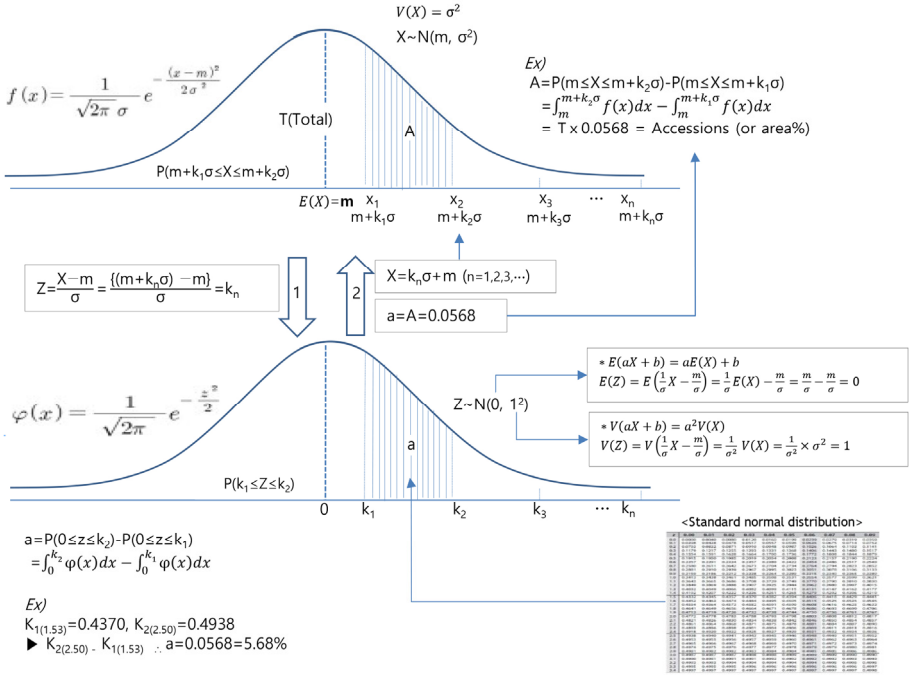
$$\text{Variability Index Value} = 1 - \sum_{i=1}^{k} P_i^2$$

정규분포의 작성



Supplement Fig. 2. Relationship between probability density histogram and normal distribution.

정규분포의 표준화



Supplement Fig. 3. The process of standardization of normal distribution.