작물육종연구 데이터분석에서 표준편차, 표준오차 및 신뢰구간의 선택과 해석

소윤섭*

충북대학교 농업생명환경대학 식물자원학과

Selection and Interpretation of Standard Deviation, Standard Error and Confidence Interval in the Data Analysis of Crop Breeding Research

Yoon-Sup So

Department of Crop Science, Chungbuk National University, Cheongju 28644, Korea

Abstract : Among agronomists, there appears to be a confusion in selecting among standard deviation (SD), standard error (SE) and confidence interval (CI) in reporting their results as figures and graphs. If there is a confusion in selection among them, there should also be difficulties in interpreting results published in peer-reviewed journals. This review paper aims to help researchers better suited for reporting their results as well as interpreting others by revisiting the definition of SD, SE and CI and explaining in plain words the concepts behind the formula. A variation among observation obtained from an experiment can be explained by the use of SD, a descriptive statistic. If one wants to draw an attention to a variation observed among plant germplasm collected from different regions or countries, SD can be reported along with the mean so that readers can get an idea how much variation exists in the particular set of germplasm. When the purpose of reporting experiment results is about inferring true mean of the population, it is advised to use SE or CI, both inferential statistics. For example, a certain chemical compound is to be quantified from plant materials, estimated mean with SD does not tell the range where the true mean content of the chemical compound would lie. It merely indicates how variable the measured values were from replications. In this case, it would be better to report the mean with SE or CI. The author recommends the use of CI over SE since CI is a sort of adjusted SE. The adjustment comes from t value that considers not only the probability but also n size.

Keywords : agricultural research, data analysis, standard deviation(SD), standard error(SE), confidence interval(CI)

서 언

농업연구에서 실험을 통해 얻어진 결과는 통계적 증거를 제시함으로써 결과의 객관성과 신뢰성을 확보한다. 실험 결과의 요약은 주로 표나 그림으로 제시된다. 표나 그림은 각 처리군의 산술평균(arithmetic mean), 범위(range), 표준편차(standard deviation, SD), 변이계수(coefficient of variation, CV), 표준오차(standard error, SE), 신뢰구간(confidence interval, CI), 그리고 유의확률(p-value)과 함께 게재된다. 일반적으로 우리가 평균이라 부르는 산술평균은 데이터의 중심경향치를 설명하는 기술통계량(descriptive statistic)이며 범위와 표준편차 등은 데이터(즉, 관측치)의 분포를 설명하는 기술통계량이다. 즉 실험자가 실험을

통해 현실화 시킨 관측치의 중심경향치와 그 관측치들이 중심경향치에 분포해 있는 산포 정도를 설명해 주는 통계량이라 할 수 있다. 하지만 표준오차는 동일한 실험을 무수히 반복적으로 수행한다고 가정하였을 때 각 실험에서 얻어지는 표본평균‘들’의 산포를 추정하는 추정통계량(inferential statistic)이다. 이러한 명백한 개념적인 차이에도 불구하고 실험자들은 논문작성 과정에서 표와 그림으로 제시하는 실험 결과의 요약에 함께 넣을 통계적 증거를 채택함에 있어 표준편차와 표준오차 사이에서 선택의 어려움을 겪고 있는 듯 하다(Altman & Bland 2005, Lee et al. 2015, Payton et al. 2000).

이러한 혼선은 농학분야 뿐 아니라 통계학적 가설검정을 실시하는 대부분의 학문분야에서 비슷한 듯 한데 의학 임상연구의 경우 표준오차의 사용은 논문의 저자가 독자들에게 연구결과의 정밀성(precision)에 관해 특별히 알리고자 하는 경우에만 한정하여 사용하도록 권고하였다(Barde & Barde 2012). 이들은

*Corresponding author (E-mail: yoonsupso@cbnu.ac.kr, Tel: +82-43-261-2517)

(Received on May 16, 2016. Accepted on May 19, 2016.)

또한 표와 그래프의 사용에 있어서도 표준오차보다는 표준편차를 우선 사용할 것과 표준편차의 사용에 있어서 “평균±표준편차” 보다는 “평균(표준편차)”와 같이 표현함으로써 95%신뢰구간과 헷갈릴 수 있는 가능성을 배제하고자 하였다.

표준편차와 표준오차의 잘못된 사용은 그 개념을 정확히 이해하지 못한 경우도 있지만 가설검정에서 대부분의 경우 귀무가설을 기각하고 대립가설이 채택된 경우에만 제대로 된 실험의 결과를 얻은 것 같은 과학계의 고정관념에서도 기인하는 듯하다. 다시 말하면 저자들이 표준편차보다 항상 작은 값을 가지는 표준오차로 그래프의 오차막대(error bar)를 표시함으로써 독자들로 하여금 관측치의 변이를 실제보다 작게 추정하도록 만들게 된다(Nagele 2003).

본 고찰에서는 통계학에서 설명하는 표준편차, 표준오차 및 신뢰구간에 대한 개념을 공식과 함께 설명하고 농업실험의 데이터분석을 통해 얻어진 결과를 표와 그래프로 요약 제시할 때 사용해야 할 바른 통계량을 가상의 농업 실험 예를 들어 제시하고자 하였다. 이는 또한 통계분석 결과 해석에 도움이 될 것이라 기대한다.

본 론

표준편차 (Standard Deviation, SD)

기술통계량인 평균과 표준편차는 실험의 관측치를 요약할 때 흔히 계산된다. 기술통계량은 어떤 정보를 기술, 즉 서술한다는 의미이다. 보통 평균이라 부르는 산술평균은 다음과 같이 정의된다(Ramsey & Schafer 2013).

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \dots\dots\dots [\text{공식1}]$$

이 기술통계량을 통해 우리는 대부분 관측치의 중심위치를 파악할 수 있다. 평균은 수많은 관측치들의 중심경향치를 제공해주는 하나 평균 하나만으로는 관측치의 분포를 알 수 없다. 따라서 우리는 관측치들의 분포경향치를 알기 위하여 표준편차를 구하며 표준편차의 정의는 다음과 같다(Ramsey & Schafer 2013).

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}} \dots\dots\dots [\text{공식2}]$$

표준편차는 각 관측치가 “평균적으로” 중심위치인 산술평균으로부터 분포한 경향치를 보여준다. 표준편차의 이러한 서술적 설명은 [공식2]로 이해하는 것보다 표준편차의 개념을 이해하는데 매우 유용하다. 산술평균을 구한 후 그 평균으로부터 각 관측치가 “평균적으로” 얼마나 떨어져 있는지 설명해 주는 정보가 표준편차이므로 표준편차를 구하기 위해서는 우선 평균에서 각 관측치의 상대적 거리를 구해야 한다. 통계학에서는 이 상대적 거리를 편차(deviation)로 정의하였다(Ramsey & Schafer 2013).

$$d_i = x_i - \bar{X} \dots\dots\dots [\text{공식3}]$$

편차들의 평균이 표준편차이므로 평균을 구하는 방식은 전체 값의 합을 합친 수의 개수로 나누는 것이므로 먼저 편차들의 합을 구해야 하는데 편차의 합은 0이므로 0을 편차의 개수로 나눌 시 그 값이 0이 되어 관측치의 분포경향치를 알 수 없다. 이를 해결하기 위하여 편차를 제곱하여 합치게 되는데 이를 자승합 또는 제곱합(sums of square, SS)이라 한다(Ramsey & Schafer 2013).

$$SS = \sum_{i=1}^n d_i^2 \dots\dots\dots [\text{공식4}]$$

소위 편차의 “평균”을 구하기 위해서 편차의 합이 자승합의 형태로 구해지면 그 값을 편차의 개수, 즉 관측치의 개수로 나누게 되는데 이 때 관측치의 개수를 n이라 한다면 자승합(SS)을 n으로 나눈 값을 분산(variance)이라 한다. 보다 정확히는 자승합(SS)을 n으로 나눈 값을 모집단의 분산(population variance, σ^2)으로 정의하며 자승합(SS)을 n-1로 나눈 값을 표본집단의 분산(sample variance, s^2)이라 한다. 이 때 n-1을 자유도(degrees of freedom, df)라 한다(Ramsey & Schafer 2013). 표본집단의 분산을 자유도로 나누어 계산하는 이유에 대해서는 본 고찰의 목적이 아니므로 생략한다.

$$s^2 = \frac{SS}{df} = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1} \dots\dots\dots [\text{공식5}]$$

이렇게 얻어진 분산은 그 단위가 원래의 단위에서 제공된 단위로 표시되는데 이를 원래의 단위로 환산하기 위하여 분산의

Fig. 1. Process of statistical inference.

제곱근을 구하게 되면 [공식2]의 표준편차가 되는 것이다. 따라서 [공식3], [공식4], [공식5]로 정의된 편차, 자승합 및 분산은 모두 [공식2]의 표준편차를 도출하는 과정에서 얻어지는 중간산물이라 할 수 있다.

만약 하나의 처리군을 $30 \pm 7\text{kg}$ (평균 $\pm$ 표준편차)로 표현했다면 대부분의 관측치는 30kg이라는 평균 주변에 위치하고 있으며 또한 관측치는 30kg을 중심으로 평균적으로 $\pm 7\text{kg}$ 의 거리 정도 떨어져 있다고 판단할 수 있다. 만약 모집단이 정규분포를 한다고 가정하면 산술평균 30kg은 중위수(median)와 동일하고 이는 곧 관측치의 50% 지점을 알려준다고 할 것이다. 이러한 가정 아래서 표준편차 7kg은 $30 \pm 7\text{kg}$ 의 지점인 23kg과 37kg 사이에 68.27%의 관측치가 위치한다는 것을 알려준다.

실험을 수행하는 대부분의 경우 실험자가 알고자 하는 전체 집단인 모집단 전체의 전수조사를 실시하기 힘들기 때문에 우리는 모집단의 일부만을 임의추출(random sampling)하여 표본집단을 생성하고 이 표본집단을 대상으로 실험을 실시한 후 관측치

를 확보한 후 이 관측치로부터 통계량(statistic)을 구하여 이 통계량을 통해 모집단의 모수(parameter)를 추정하게 된다. 따라서 대부분의 경우 우리는 모집단의 분산을 구하는 것이 아니라 표본집단의 분산을 구한다. 즉 우리가 일반적으로 이야기 하는 표준편차는 대부분이 표본집단의 표준편차인 것이다.

표준오차 (Standard Error, SE)

표준편차가 실제 수행한 실험의 결과 얻어진 관측치의 분포에 관한 정보를 제공해 주는 것이라면 표준오차는 한 번만 수행한 실험을 동일하게 무수히 반복해서 수행하였을 때를 가정하고 얻어지는 정보이다. 만약 실제로 수행한 실험(realized experiment)과 동일한 개체수를 가지는 표본집단을 그 모집단에서 반복적으로 생성하고 그 때 마다 표본집단의 평균을 구했다고 가정하자(Fig. 1). 표본을 추출함으로써 인해 발생하는 오차로 인해 표본평균은 변하지 않는 모수인 모평균과 다른 값을 가질 수 있으며 표본평균들간에 발생한 변이로 인해 표본평균들의 평균과 표준편차를

계산할 수 있다. 이 “표본평균들의 표준편차”를 표준오차라고 하는데 실제 수행한 실험이 아니기 때문에 표본평균들의 표준편차는 실제로 구할 수 없는 듯 보이나 한 번의 실험을 통해서도 아래의 [공식6]에 의해 표준오차를 계산할 수 있다(Ramsey & Schafer 2013).

$$S_{\bar{X}} = \frac{s}{\sqrt{n}} \dots \dots \dots \text{[공식6]}$$

표준오차의 기호는 주로 $S_{\bar{X}}$ 로 표시하는데 S 는 표준편차를 의미하며 $\bar{X}$ 는 표본평균을 의미하므로 표준오차는 표본평균들의 표준편차라는 개념을 잘 설명해 주는 기호라 할 수 있을 것이다. 표준오차는 [공식6]에서 보듯이 표준편차에 개체수의 제곱근을 나누어 준 값으로 계산된다. 따라서 한 실험에서의 표준오차는 항상 표준편차보다 작은 값을 가지며 이러한 이유로 그래프로 실험결과를 요약할 시 그래프의 목적과 상관없이 표준편차보다 표준오차로 오차막대를 표시하여 측정값의 변이가 작게 보이도록 유도할 가능성이 있다.

하나의 모집단에서 표본집단을 반복적으로 생성하고 그 때마다 표본집단의 평균을 구하였다면 그 표본평균들의 분포는 모집단의 분포와 상관없이 정규분포를 따르는데 이를 확률이론에서는 중심극한정리(central limit theorem)라 한다. 표본평균의 수가 무한대에 수렴한다면 그 표본평균들의 분포는 정규분포에 더 근접하게 되며 정규분포에서 평균 $\pm$ 표준편차는 68.27%의 표본평균들을 포함한다. 이 표본평균들의 분포에서 표준편차는 표준오차이므로 만약 한 처리군을 $30 \pm 7\text{kg}$ (표본평균 $\pm$ 표준오차)로 표현하였다면 동일한 실험을 반복적으로 무한대로 수행하였을 때 표본의 평균은 대략 23kg과 37kg의 범위 사이에서 약 68.27%의 표본평균이 위치할 것이라는 것을 짐작할 수 있다.

우리는 모집단의 전수조사를 할 수 없기 때문에 모집단과 관련된 모수인 모평균(population mean, μ)도 알 수 없는 상수로 규정하고 유한한 개체수를 가지는 표본집단을 모집단으로부터 임의추출하여 표본집단으로부터 통계량인 표본평균($\bar{X}$)을 구하여 이를 통해 모평균을 추정하게 된다. 문제는 표본을 추출할 때 마다 표본평균이 표본추출오차로 인해 모평균과 달라지기 때문에 특정 비율(68.27%)의 표본평균들이 속할 범위를 표준오차라는 통계량으로 제시함으로써 알 수 없는 모수인 모평균의 진정한 위치를 추정하게 되는 것이다. 따라서 이미 수행한 실험에서 얻은 관측치의 분포를 서술하는 기술통계량인 표준편차와는

달리, 비록 관측치로부터 구할 수 있는 값이긴 하지만, 표준오차는 모평균의 진정한 위치, 즉 값을 추정하기 위한 정보를 제공해 줌으로 추정통계량이 된다.

표준오차와 표준편차에서의 “표준(standard)”이라 함은 표준화된(standardized) 오차와 편차라는 것이다. 통계학에서 표준화는 대체로 데이터를 산술평균으로 빼고 그 값을 표준편차로 나누어 주는 것을 의미한다. 정규분포를 평균이 0이고 표준편차가 1인 표준정규분포로 변환하는 과정도 전체 데이터에서 평균을 빼고 표준편차로 나누어 주는 표준화를 거치는 것이다. 만약 정규분포하는 모집단에서 표본을 추출하여 표본의 평균을 구하였다면 이 표본의 평균에서 모평균을 빼고 표본평균들의 표준편차인 표준오차로 나누어 주면 표본평균들의 분포를 표준화시킬 수 있다. 이것이 t 검정통계량 계산시 분모에 표준편차 대신 “표본평균들의 표준편차”인 표준오차를 사용하는 이유이다.

그런데 표준오차는 사실상 많은 정보를 제공하지 않고 단지 신뢰구간(confidence interval, CI)를 계산하기 위한 중간 과정으로 이용된다.

신뢰구간(Confidence Interval, CI)

표본평균의 신뢰구간은 알려지지 않은 상수인 모평균이 속해 있을 것이라 예상되는 값의 범위라 할 수 있으며 다음의 공식으로 구할 수 있다(Ramsey & Schafer 2013).

$$\bar{X} \pm t_{\alpha(df)} S_{\bar{X}} \dots \dots \dots \text{[공식7]}$$

여기서 $t_{\alpha(df)}$ 는 유의수준 α 와 자유도가 df 인 t분포에서의 기각값을 의미한다. 유의수준을 0.05로 할 때 신뢰구간은 $(1-\alpha)*100=95\%$ 의 신뢰수준을 가지게 되며 만약 95%의 신뢰구간을 $30 \pm 7\text{kg}$ 로 표시하였다면 95%의 확률로 그 범위인 23~37kg이 알 수 없는 모평균을 포함할 것이라는 추정이다. 모수는 변하지 않는 상수이므로 고정된 위치를 가지고 있으며 따라서 신뢰구간의 해석 시 95%의 확률로 모평균이 신뢰구간에 포함'된다'고 표현하는 것은 잘못된 표현이라 할 수 있다.

Fig. 2는 유의수준 0.05에서 자유도에 따른 기각값을 나타낸 그림이다. 자유도가 5이하에서는 기각값이 상대적으로 매우 크며 자유도가 5 이상부터 기각값이 2에 수렴하는 것을 볼 수 있다. 통계학에서 자유도는 표본집단의 개체수인 n 과 밀접한 관계를 가지고 있으므로 표본의 개체수가 증가하면 기각값이 작아지며 결국 신뢰구간은 좁아지게 되며 이는 모수추정치의

Fig. 2. The figure shows that the t value at certain level of significance is a function of the degrees of freedom.

정밀성(precision)이 증가함을 의미한다. 자유도가 1, 즉 $n=2$ 일 때의 기각값은 약 12.7이며 자유도가 2 ($n=3$)일 때 기각값은 약 4.3으로 기각값이 3배정도 줄어들며 자유도가 9 ($n=10$) 일때 기각값이 2와 가까워지며 그 이상에서는 2와 큰 차이가 없으며 $df=2$ 일 때 보다 2배정도 줄어드는 것을 볼 수 있다. 따라서 만약 그래프의 오차막대가 표준오차로 표시되었다면 $n=3$ 일 때는 표준오차에 4를 곱하고 $n=10$ 이상에서는 표준오차에 2를 곱함으로써 대략의 95% 신뢰구간을 파악할 수 있다.

포장실험에서의 경우 각 표본집단은 각 처리군이 될 것이며 각 처리군의 n 은 대부분의 경우 반복의 수인 r 값과 동일하다 할 수 있다. 모수추정치의 정밀성을 높이기 위해서 반복의 수를 10으로 하기는 현실적으로 힘들겠지만 포장실험에서 가능하면 2반복으로 실험하는 것보다 3반복으로 실험하는 것이 정밀성을 3배정도 높일 수 있다는 것을 그림2는 보여준다.

적 용

Fig. 3은 가상의 데이터를 이용하여 분산분석을 실시하고 각 처리군의 평균을 막대그래프로 표시한 것이다. 처리군은 3개였으며 각 처리군당 개체수는 각각 5, 13, 그리고 10개였다. Fig. 3(a)는 각 처리의 평균과 “표준편차”를 나타낸 것이다. 표준편차는 각 데이터의 변이분포를 나타내는 것이므로 정규분포를 한다는 가정하에 오차막대는 각 처리군에 속하는 관측치들의 약 68%가 오차막대 내에 위치한다는 것을 알려준다. 처리군 A의 오차막대, 즉 표준편차가 가장 크므로 데이터의 산포가 가장 넓고 처리군 B의 오차막대가 가장 짧으므로 그 산포가

가장 좁은 것을 나타낸다. 만약 처리군(예 품종)에 따른 특정 성분함량의 차이를 파악하기 위해 처리군당 동일한 수의 시료를 채취하여 HPLC등과 같은 분석기기를 통해 분석한 데이터를 통해 각 처리군의 표준편차를 표시하였다면 각 처리군의 표준편차의 크기를 비교함으로써 측정시 오류를 판별할 수 있을 것이다. 즉 특정 처리군에서 다른 처리군에 비해 매우 큰 표준편차값을 보인다면 이것은 대부분 측정 오류라 할 수 있다. 일반적으로 이런 경우 QQ plot등을 통해 특이값 제거 등의 과정을 거쳐야 할 것이라 생각된다. 육종학에서 오차막대를 표준편차로 사용하는 경우는 보통 품종의 육종시기별 특정 형질의 비교나 유전자원 수집지에 따른 변이 분포를 파악하기 위해 실험을 실시하는 경우라 할 수 있을 것이다. 이 경우 각 수집지 자원들의 형질 평균이 어디에 위치하며 이들의 특정(약 68%)수가 평균에서 어느 정도의 범위에 분포하는 것인가도 매우 중요한 정보라 할 수 있을 것이다.

Fig. 3(b)는 각 처리의 평균을 “표준오차”와 함께 나타낸 것이다. 이 경우 (a)와는 달리 이러한 실험을 반복적으로 무수히 수행할 시 각 처리 평균의 추정치가 대략적으로 위치 할 범위를 알려준다. 오차막대가 (a)보다 좁은 범위를 나타내는 것은 표준편차를 개체수의 제곱근으로 나누어 계산하였기 때문이다. 우리는 흔히 (a) 또는 (b)로 오차막대를 표시하는데 (a)의 경우 그 정보가 분명한데 반하여 (b)의 경우는 추정치의 대략적 범위 이외의 유의성 검정 결과 등에 대해 추론하기에는 미흡하다. 즉 두 처리의 평균과 표준오차를 육안으로 확인함으로써 오차막대가 겹치지 않는다고 하여 두 처리군간의 유의성을 확신할 수는

Fig. 3. Treatment means are displayed with various statistics. (a) mean with standard deviation, (b) mean with standard error of each treatment group, (c) mean with 95% confidence interval computed from the t-value and standard error associated to each treatment group, (d) mean with 95% confidence interval from analysis of variance.

없는 것이다. 따라서 유의성검정 결과와 표본평균의 추정치의 범위를 보다 명확히 전달하기 위해서는 표준편차보다는 신뢰구간이 더 좋은 대안이 될 수 있다.

Figure 3(c)와 (d)는 평균을 95%신뢰구간으로 표시한 막대그래프이다. (c)의 경우 각 신뢰구간의 계산은 다음과 같이 실시되었다. 처리군 A로 예를 들면 처리군 A의 개체수는 $n=5$, 표준편차는 $s=4.64$ 였으며 표준오차는 $s/\sqrt{n}=2.07$ 이었다. 자유도 $df=4$ 이므로 유의수준 $\alpha=0.05$ 에서 $t_{0.05(4)}=2.78$ 이었고 이를 표준오차와 곱하였다($t_{0.05(4)} \times s/\sqrt{n}=5.76$). 이를 평균 $\pm$ 95% 신뢰구간으로 처리군 A의 막대그래프에 표시하였다. 각 처리군에서의 표준오차와 t값의 사용은 특별히 문제가 없는 것처럼 보인다. 하지만 여기에는 한가지 큰 오류가 포함된다.

보통 이러한 실험의 목적은 각 표본평균간 유의한 차이가 있는지에 대한 가설검정을 위한 것인데 이를 위해 우리는 모든 표본평균에는 차이가 없다는 귀무가설 하에서 분산분석을 실시한다. 표본평균에 차이가 없다는 것은 엄밀히 말해 각 표본평균이 추정하는 모평균간에 차이가 없다는 것이며 등분산성(homoscedasticity) 하에서 이것은 모든 표본집단, 즉 처리군은 동일한 모수를 가지는 하나의

모집단으로부터 유래된 것이란 가정이다. 그러므로 각 처리군에서 추정된 표준편차는 각 처리내의 관측치만을 가지고 추정된 통계량인므로 귀무가설 하에서 모집단의 표준편차를 가장 정확히 추정하는 통계량은 실험 전체 관측치를 모두 이용하여 얻어진 표준편차가 될 것이다. 이를 위하여 각 처리군에서 얻어진 분산에 대해 각 처리군의 자유도를 통해 가중치를 적용한 합동분산(pooled variance, S_p)을 구하게 되는데 두 평균의 비교에서는 합동분산을 아래 [공식8] 같이 각 처리군 분산의 가중평균으로 구하게 되며 여러 평균의 비교에서는 분산분석에서 얻어진 오차의 평균자승합(MS_{error})값이 합동분산 값에 해당한다.

$$S_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2} \quad \dots\dots\dots \text{[공식8]}$$

Fig. 3(d)는 분산분석을 통해 획득한 오차의 평균자승합을 이용하여 구해진 95% 신뢰구간을 표시한 것이다. 처리군 A만을 예로 들면 개체수는 $n=5$, 표준편차는 $s=4.64$ 였으나 평균자승합

Fig. 4. SAS code and output for obtaining 95% confidence interval.

11.89로부터 얻어진 표준편차인 3.45를 대신 이용하여 표준오차를 구하였다($\sqrt{MS_{error}} / \sqrt{n} = 1.54$). 또한 (c)에서는 $df=4$ 일 때의 $t_{0.05(4)}=2.78$ 값을 이용하지만 귀무가설하에서 얻어진 평균 자승합으로 처리군 A 내의 분산값을 대체한 것처럼 t값 또한 오차의 자유도 $df=(5-1)+(13-1)+(10-1)=25$ 일 때의 $t_{0.05(25)}=2.06$ 값으로 대체하여 95%의 신뢰구간을 확보하게 된다 ($t_{0.05(25)} * (\sqrt{MS_{error}} / \sqrt{n}) = 3.17$).

따라서 처리군 A의 95% 신뢰구간이 (c)에 비해 (d)가 작은 것은 (c)에서는 처리군 A의 $n=5$ 로부터 확보한 표준오차와 t값을 사용하는데 반해 (d)에서는 전체 관측치로부터 얻어진 표준오차와 t값을 사용하였기 때문이며 95%의 신뢰구간의 폭이 더 좁다는 것은 그만큼 추정치의 정밀도가 더 높다는 것을 의미한다.

분산분석을 통해 귀무가설이 기각되고 대립가설이 채택되었다면 두 평균간 비교를 통해 통계적 유의성이 인정되는 차이를 가지는 처리군을 파악하는데 (d)에서는 처리군들간의 신뢰구간이 중첩되는지 그렇지 않는지에 따라 대략적인 유의성 검정이 가능해진다. 만약 두 처리군의 신뢰구간이 중첩되지 않는다면 신뢰구간을 표시할 때 설정한 유의수준에서 유의성이 인정될 가능성이 높으며 그렇지 않다면 그 가능성이 매우 낮을 것이다. 실제로 최소유의차검정 결과 두 오차막대가 중첩이 되는 처리군

B와 C는 유의성이 인정되지 않았고 두 오차막대가 중첩되지 않는 처리군 A와 B, 처리군 A와 C에서는 유의차가 인정되었다.

여기서 “정확한”이 아니라 “대략적인” 유의성 검정이 가능해진다는 이유는 분산분석 이후의 post-hoc 과정인 여러 평균들의 다중검정에는 제1종의 오류 제어를 위해 다양한 방법이 사용되기 때문이다. 최소유의차검정, 던컨의 다중검정(duncan’s multiple range test), tukey, scheffe 등과 같은 다양한 검정법마다 추정된 두 평균의 차이와 연관된 95%의 신뢰구간이 달라지기 때문에 그래프에서 두 평균 간의 차이를 막대그래프로 보여주는 것이 아니라 우리가 일반적으로 이용하는 각 처리군의 평균으로 표시할 경우 평균의 다중비교에 사용한 방법에 맞는 95%의 신뢰구간을 정확히 표시하는 것은 몇 번의 계산을 더 해야만 하는 번거로운 작업이 될 것이다.

Fig. 4는 통계프로그램인 SAS의 GLM procedure를 이용하여 분산분석을 실시 후 means statement를 이용하여 그림에 오차막대를 그리기 위해 필요한 통계량을 구하는 과정을 나타낸 것이다. 여기서는 분산분석을 통해 이미 처리군들 간의 유의성이 인정된 후의 과정만을 다루었으며 class와 model statement의 경우 요인수, 시험구 배치법에 따라 달라질 수 있다. 분산분석을 통해 유의성이 인정되면 처리 평균들간의 다중검정을 실시한다.

예에서는 means statement와 t 옵션을 이용하여 최소유의차검정을 실시하였다(Fig. 4(a)). 그 결과는 각 처리 평균 간의 차이와 그 차이의 95% 신뢰구간 그리고 차이의 유의성 유무가 별표로 표시된다. 하지만 우리가 원하는 각 처리군의 평균, 표준편차는 나타나지 않는다. 따라서 이번에는 means statement에서 t 옵션을 삭제하고 다시 실행을 하거나 clm 옵션을 넣어 실행을 하면 각 처리의 평균과 표준편차를 얻을 수 있다. 또한 means statement에 t 옵션과 clm 옵션을 함께 사용하면 t 옵션만을 사용했을 때의 다중비교 결과는 얻을 수 없으나 각 처리군의 평균과 95% 신뢰구간을 확보할 수 있다. SAS는 유의수준의 초기값으로 0.05를 사용한다. 만약 유의수준 0.01에서 99%의 신뢰구간을 구하고자 한다면 means statement에 alpha=0.01의 옵션을 넣어 구할 수 있다. SAS를 통해 얻어진 95%의 신뢰구간은 앞서 언급한 Fig. 4(d)의 신뢰구간과 동일하다. 또한 신뢰구간은 표준오차를 t값으로 보정한 일종의 오차막대라 생각할 수 있기 때문에 n값에 상관없이 서로 간의 비교가 가능한 장점이 있다(Cumming et al. 2007).

지역적응성 비교시험, 시비수준 결정을 위한 포장 시험 등에서는 관측치의 변이 분포 보다는 수량 추정치의 추정범위가 더 중요한 정보가 되므로 이러한 경우 표준편차보다는 표준오차나 더 나아가 신뢰구간으로 표시하는 것이 더 바람직할 것이라 생각되며 표준오차와 신뢰구간 중에서는 신뢰구간이 보다 명확한 정보를 전달할 것이다.

결 론

농업실험 데이터의 분석을 통해 얻어진 통계량 중 산술평균만으로는 관측치의 중요한 정보를 전달하기에는 부족하다. 따라서 표준편차, 표준오차, 신뢰구간 등과 같은 통계량을 함께 전달함으로써 모집단의 특성에 대한 보다 정확한 추론이 가능하다. 하지만 연구자들은 표준편차, 표준오차, 신뢰구간의 선택과 해석에서 어려움을 겪는 듯 하다. 본 고찰은 이러한 선택의 어려움을 해소하고 학술논문에서 제시된 결과에 대한 바른 해석을 위하여 표준편차, 표준오차 및 신뢰구간에 대한 개념적 설명을 정의와 공식을 병행하여 설명함으로써 연구자들의 이러한 고충을 해결하고 힘든 실험을 통해 확보한 데이터의 질을 향상시키기 위해 작성되었다.

본 고찰을 작성하기 위하여 국내 농업관련 학회지의 최근호를 조사하였는데 평균 이외 표준편차, 표준오차 또는 신뢰구간을 전혀 표시하지 않는 논문들도 많았으며 그래프의 오차막대가

표준편차인지 표준오차인지 표시하지 않은 논문들도 다수였다. 따라서 논문의 신뢰성을 높이고 결과에 대한 바른 해석을 위해서도 경우에 맞는 오차막대를 선택하여 사용하여야 할 것이다.

관측치의 변이 분포 정도를 전달하고자 한다면 기술통계량인 표준편차를 표시하여야 할 것이고 그래프와 표가 유의성 검정 결과 전달에 보다 큰 목적이 있다면 추정통계량인 표준오차 또는 신뢰구간을 사용해야 할 것이다. 이 중 신뢰구간이 결과에 대한 정보를 보다 정확히 전달해 주는 통계량이므로 신뢰구간의 선택을 추천한다. 그래프와 표에는 반드시 n을 표기해야 할 것이며 그래프에 오차막대가 표시되었다면 어떤 통계량이 사용되었는지에 대한 명확한 표시가 있어야 할 것이다.

또한 일반적인 농업실험의 경우 표준편차, 표준오차, 신뢰구간은 모두 관측치의 정규분포성에 기반함으로 데이터 분석 전 QQ plot 등을 통한 오차의 정규성 조사와 필요시 box-cox power transformation 등의 방법을 통한 데이터의 수치변환을 하여 정규성을 확보하는 것이 논문에서 제시한 표준편차, 표준오차 및 신뢰구간의 유효성을 해치지 않는다는 점을 상기해야 할 것이다.

적 요

표준편차는 가상이 아닌 실제 실험에서 얻어진 관측치의 분포에 관한 기술통계량으로 관측치가 “평균적”으로 산술평균에서 떨어진 산포 정도를 표시하는 값이다. 모집단이 정규분포를 할 때 평균±표준편차는 약 68%, 평균±2표준편차는 관측치의 약 95%의 관측치를 포함하는 범위이다. 따라서 관측치의 변이 분포 전달이 표나 그림의 중요 목적이면 표준편차를 표시한다.

표준오차는 동일한 실험을 반복적으로 무수히 실시하는 가상의 경우 매 번의 시행에서 얻을 수 있는 표본집단의 평균들의 분포와 관련된 추정통계량이다. 이러한 가정 하에서 얻어진 표본평균들의 표준편차를 표준오차라 하며 한 번의 실험을 통해 구할 수 있다.

신뢰구간은 표준오차를 t값으로 보정한 값이며 유의수준을 고려한 값이므로 표준오차보다 보다 명확한 추정 범위를 제공한다. 실험을 통한 추정치 확보가 목적이라면 표나 그림에 표준오차나 신뢰구간을 표시한다.

사 사

이 논문은 2014년도 충북대학교 학술연구지원사업의 연구비 지원에 의하여 연구되었음.

REFERENCES

1. Altman DG, Bland JM. 2005. Standard deviations and standard errors. *BMJ* 331: 903.
2. Barde MP, Barde PJ. 2012. What to use to express the variability of data: Standard deviation or standard error of mean. *Perspect. Clin. Res.* 3: 113-116.
3. Cumming G, Fidler F, Vaux DL. 2007. Error bars in experimental biology. *J. Cell Biol.* 177: 7-11.
4. Lee DK, In J, Lee S. 2015. Standard deviation and standard error of the mean. *Korean J. Anesthesiol.* 68: 220-223.
5. Nagele P. 2003. Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *Br. J. Anaesth.* 90: 514-516.
6. Payton ME, Miller AE, Raun WR. 2000. Testing statistical hypotheses using standard error bars and confidence intervals. *Commun. Soil Sci. Plant Anal.* 31: 547-551.
7. Ramsey FL, Schafer DW. 2013. *The statistical sleuth: a course in methods of data analysis*. 3rd edition. Brooks/Cole Cengage Learning, Australia.