

# 수의학 관련 자료에 대한 통계분석 기법

강원대학교 수의학부대학

박 선 일

자료 분석을 위해 논의해야 할 내용이 매우 방대하기 때문에 자세한 내용은 강좌에서 다루기로 하고 학회지에 발표되는 논문에서 흔히 범하는 오류와 관련하여 몇가지 중요한 통계적 개념을 소개한다.

## 1. 실험설계와 자료 분석에서 고려할 사항

실험을 계획할 때 고려해야 할 사항과 자료 분석 절차에 대한 이해를 돕기 위하여 간단한 예를 들어보자. 사료 첨가제의 효과를 평가하는 연구에서 실험대상 동물을 세 군으로 분류하여 제1군에는 첨가제 농도를 5%로 투여하고, 제2군에는 농도 2% 투여, 세 번째 군에는 대조군으로 첨가제를 투여하지 않은 상태로 실험하여 종료시점에서 각 군별 체중을 측정하여 증체율 혹은 평균을 계산함으로써 사료첨가제의 효과 여부를 판단하게 된다. 이와 같은 실험에서 고려해야 할 사항은 다음과 같다.

첫째, 두 군에 할당된 실험동물이 결과(증체율)에 영향을 미칠 수 있는 조건이 모두 동일하게 분포하고 있는지를 검토해야 한다. 모든 실험동물이 임상적으로 건강하다고 할지라도 개체의 유전적 특성 등에 기인한 고유한 차이로 두 군 간 차이가 우연히 발생할 수 있기 때문이다. 실험연구(experimental study)는 연구자가 통제할 수 있는 환경에서 수행되므로 결과에 영향을 미칠 것으로 생각되는 요인이 있다면 블록화 실험계획(block design)을 사용해야 한다.

둘째, 증체율의 변동성과 관련하여 체중을 1회 측정할 것인지 아니면 측정횟수를 늘릴 것인지를 결정해야 한다. 동일한 개체에서 체중을 2회 이상 반복측정하는 경우 분석방법이 달라진다.

셋째, 실험결과는 증체율(비율 혹은 평균)이므로 이러한 변수를 분석하는데 적용할 분석기법을 검토해야 한다. 요인의 개수, 요인 수준의 개수, 종속변수의 척도 등에 따라 분석기법은 달라진다.

넷째, 측정된 체중 자료가 정규분포를 만족하는지를 검토해야 한다. 대부분의 모수분석(parametric analysis)은 정규분포 가정을 근거로 하기 때문에 정규성을 만족하지 못하면 비모수분석(non-parametric analysis)을 고려해야 한다.

다섯째, 유의수준으로 흔히 5%를 사용하지만 연구목적과 내용에 따라서는 10%나 1% 등으로 조정할 필요가 있다.

여섯째, 실험군간 차이가 있을 경우 특정한 처리군 간의 차이를 검증하기 위하여 다중비교(multiple comparison)를 실시하게 된다. 연구의 구체적인 목적이 모든 처리군 간 비교인지 아니면 대조군과 나머지 군 간의 비교인지에 따라 다중비교 방법이 다르고 경우에 따라서는 대비(contrast)를 이용하여 수행할 수 있다.

일곱째, 실험 자체가 매우 많은 비용을 필요로 하는 상황에서는 실험동물의 수 즉 표본크기를 연구계획 단계에서 조정해야 하며 연구의 목적을 달성하는데 필요한 최소한의 동물을 대상으로 하는 것이 중요하다. 표본

크기가 부족하면 실제로 존재하는 중요한 차이를 검출하지 못할 가능성이 높아지며 반대로 표본크기가 너무 많으면 중요치 않은 작은 차이도 통계학적으로 유의한 차이가 있다는 판단하게 된다.

여덟째, 검정의 형태를 단측 (one-tailed test) 혹은 양측검정 (two-tailed test)으로 수행할 것인지를 판단한다. 양측검정은 두 군간 차이의 존재유무만을 평가하는 것이고 단측검정은 이러한 차이가 분포의 위 혹은 아래 어떠한 방향성 (direction)을 가지고 있는지를 검증하는 것으로 표본크기와 관련이 있으며 귀무가설의 수용여부를 판단하는 기각역의 판단기준이 달라진다.

아홉째, 분석의 최종단계에서 처리군 간의 차이를 흔히 p-value로 평가하는데 실험결과에서 나타난 두 군 간 차이에 대한 통계학적 유의성이 반드시 임상적으로 중요하다는 것을 의미하는 것은 아니다. 이를테면 두 군 간 증체율 1%의 차이가 설명 통계학적으로는 유의하지만 임상적으로는 큰 차이가 없을 수 있다는 점이다. 또한 유의확률을 표현할 때  $p < 0.05$ 로 표현하는 것 보다는 패키지에서 출력된 결과를 그대로 인용하는 것이 바람직하며 최근에는 신뢰구간으로 요약할 것을 적극 권장하고 있는 추세다.

## 2. 자료의 형태

의학 분야에서 접하는 자료의 형태는 크게 네가지로 구분된다. 명목 (nominal)척도는 예를 들어 고양이 혈액형, 환자의 성별 (gender), 품종 (breed) 등과 같이 수학적 연산이 불가능한 척도다. 순위 혹은 서열 (ordinal) 척도는 명목척도에 순서의 개념이 추가된 자료로 순위간의 차이가 불분명하고 알려지지 않은 경우가 대부분이기 때문에 여전히 정량적 계산을 할 수 없는 형태다. 환자의 통증을 1-5 단계로 평가한 자료나 dipstick을 이용한 요분석 (urinalysis) 결과를 +, ++, +++ 등은 전형적인 순위형 척도다. 구간 (interval)과 비척도 (ratio) 자료는 이웃하는 수치간의 차이가 일정하기 때문에 모든 산술적인 연산이 가능한 자료다. 구간척도는 진정한 0의 값을 갖지 못한다는 점이 비척도와 다르고 온도는 대표적인 예다. 척도와 관련하여 중요한 것은 가능한 상위척도로 자료를 수집해야 한다는 점이다. 비척도는 필요시 순위나 명목척도로 변환이 가능하지만 명목척도나 순위척도는 상위의 척도로 변환이 불가능하기 때문이다. 또한 분석단계에서 상위척도를 하위척도로 변환하게 되면 정보의 손실이 초래되기 때문에 가능한 해당 척도에 부합하는 통계기법을 사용하는 것이 최선의 방법이다.

## 3. 자료요약 방법

중앙집중성: 자료를 요약하는 중요한 두가지 속성은 개별 관찰치들이 평균적으로 분포하는 위치 (location) 혹은 중앙집중성 (central tendency)과 자료들이 평균 위치로부터 얼마나 근접해 있는지에 대한 변동성 (variability) 혹은 산포성 (spread, dispersion)이다. 이러한 두 속성은 자료의 척도에 따라 서로 다른 방법으로 요약된다. 명목척도는 최빈값으로 요약하는 반면 순위척도는 중위수 (median, 관찰치 오름차순으로 정렬할 때 중앙에 위치하는 값으로 관찰치의 개수가 짝수이면 중앙에 위치하는 2개 관찰치의 평균값)를 사용한다. 자료의 분포가 대칭이고 이상점 (outlier)이 없는 연속형 자료는 산술평균 (mean)을 사용하며 자료의 분포가 비대칭이거나 이상점이 있는 경우에는 중앙값을 사용한다. 명목척도를 갖는 자료는 관찰빈도를 이용하여 백분율로 정리하며 순위척도는 중위수나 백분율을 사용하고 구간척도 이상의 자료는 평균이나 중위수를 사용

한다. 산술평균값은 이상점에 큰 영향을 받지만 중위수는 영향을 받지 않기 때문에 매우 크거나 작은 수치가 있는 자료를 요약할 때 유용하다.

산포성: 위치모수와 더불어 분포의 형태를 좌우하는 중요한 지표인 산포성은 표준편차 (standard deviation), 범위 (range, 최대값과 최소값의 차이로 단일 수치이지만 대부분의 논문에서 최소값과 최대값을 각각 표기함), 사분위수 범위 (interquartile range, 25% 백분위수와 75% 백분위수의 차이) 등을 사용한다. 범위는 이상점이 있을 경우 전체 자료를 대표하는 수단으로 유용하지 못하고 최소와 최대값 사이에 위치하는 수치에 대한 아무런 정보를 제공하지 못하는 단점이 있다. 중위수와 마찬가지로 사분위범위는 이상점에 영향을 받지 않기 때문에 대칭분포를 보이지 않는 자료를 요약할 때 사용한다.

#### 4. 자료변환

연구자가 관심을 두고 측정된 모든 자료는 연구 대상이나 방법 등에 따라 무수히 많은 고유의 분포가 존재하지만 대부분 단봉형 (unimodal) 분포를 보이는데 이는 관찰치의 평균위치에 최고치 (modal)와 분포의 양쪽으로 꼬리를 갖는 대칭형 분포라는 것을 의미한다. 그러나 모든 단봉형 분포가 대칭인 것은 아니고 왼쪽이나 오른쪽으로 긴 꼬리를 갖는 비대칭형 (asymmetrical) 분포를 보이기도 한다. 이러한 왜곡 (skewness)의 형태는 오른쪽으로 긴 꼬리를 이룰 때 양성왜곡 (positive, right skewness)이라고 하며 이는 대부분의 관찰치들이 상대적으로 작은 값을 갖지만 일부 관찰치는 매우 큰 값을 갖고 있는 자료라는 것을 시사한다. 양성왜곡을 보이는 자료에서 분포의 오른쪽에 위치하는 이상점들이 평균에 영향을 주기 때문에 중위수는 항상 평균보다 더 작다. 한편 왼쪽으로 긴 꼬리를 이루는 자료를 음성왜곡 (negative, left skewness)이라고 하며 이 경우 중위수는 평균보다 항상 크다. 이와 같이 왜곡된 분포나 특히 도수로 측정되는 자료는 정규분포를 따르는 예가 매우 드물고 심지어 연속형으로 측정된 자료도 정규분포를 따르지 않는 경우도 많다. 이 경우 자료에 대한 요약통계량으로 평균은 적절하지 못하기 때문에 수학적 방법으로 정규성을 만족시키는 방법을 고려해야 한다. 자료변환 (transformation)은 원자료를 분석에 적합하도록 자료 고유의 특성을 변형시키는 것을 의미한다. 통계적 추론이 가능하기 위해서 흔히 정규분포를 가정하지만 이러한 가정을 충족하지 않는 자료가 많기 때문에 비정규분포 자료를 정규분포 자료로 변환하는 것이다. 특히 양성왜곡을 보이는 자료를 제곱근 (square root), 대수 (logarithm) 혹은 역수 (inverse) 등과 같은 자료변환을 시도하면 평균이 요약통계량으로 적절한 경우가 많다.

#### 5. 표준오차

연구자가 실험대상으로 선정한 표본은 상위 모집단에서 추출 가능한 모든 표본 중 어느 한 표본에 불과하다. 이는 모집단에서 확률적으로 선발된 다양한 표본들에 대하여 평균을 계산한다면 서로 다른 값을 보일 것이고 하나의 표본에서 얻은 통계량이 반드시 모집단의 특성을 대표한다고 말하기 어렵다는 것을 의미한다. 따라서 표본 추정치에 내재된 변동성을 검토하는 것이 중요하다. 모집단에서 동일한 표본크기 (n)를 갖는 모든 가능한 표본을 선발한다고 가정하면 각각의 표본들은 서로 다른 평균과 표준편차를 갖기 때문에 분포도 다를 것



이다. 표준편차는 이러한 단일 표본에 대하여 평균과 개별 관찰치 간의 변동성 즉 편차를 측정하는 수단이지만 표준오차 (standard error of the mean, SE, SEM)는 단일 표본에서 모집단 평균의 참값을 추정할 때 추정치의 불확실성 (uncertainty)에 대한 정밀도 (precision)를 나타내는 용어로 표본평균들의 표준편차를 의미하므로 자료의 기술통계량으로 사용하는 것은 적절하지 못하다.

## 6. 유의수준과 유의확률

통계검정에서 흔히 사용하는 유의수준 (significance level)은 귀무가설을 기각할 때 이를테면 두 군의 평균에 차이가 있다는 결론을 내릴 때 범할 수 있는 오류 (error)의 크기를 연구자가 어느 정도까지 허용할 것인지를 설정하는 수준이다. 예를 들어 5%의 유의수준에서 검정하였다다는 것은 검정결과 귀무가설을 기각하고 새로운 치료법의 효과가 기존 치료법의 치료효과보다 우수하다는 결론을 내릴 때 이러한 결론이 거짓일 확률로 5%는 허용한다는 것을 선언하는 것이다. 통계용어로 표현하면 참인 귀무가설을 기각하는 제1종 오류의 최대 허용수준이다. 가설검정에서 발생하는 와 오류를 설정하는 절대적인 기준은 없고 의학연구에서는 제1종 오류의 크기를 0.05, 0.01, 제2종 오류의 크기를 0.1, 0.2를 흔히 사용한다. 연구자는 치료효과에 차이가 있다는 긍정적인 결과 (positive result)를 주장하고 싶은 욕구가 있기 때문에 통계검정에서 그러한 결론을 내릴 때 범할 수 있는 유의수준을 연구 상황과 목적에 따라 적절히 결정할 필요가 있다.

한편 유의확률 (significance probability)은 귀무가설이 참이라는 가정에서 얻은 검정통계량보다 더 극단의 통계량을 얻을 확률로 귀무가설을 기각하는 유의수준의 최소값이다. 즉 두 군간 관찰된 차이가 순전히 우연에 의하여 나타날 확률이 어느 정도인지를 측정하는 수단으로 귀무가설이 참일 때 순전히 우연에 의하여 관찰된 차이 (검정통계량)와 같거나 더 큰 차이를 보게 될 확률이다. 유의확률은 확률의 개념이므로 0에서 1까지의 값을 갖는데  $p=0$ 이라는 것은 관찰된 차이가 우연에 기인하였을 가능성이 없음을 의미하고,  $p=1$ 은 관찰된 차이가 순전히 우연에 의한 것이므로 두 군간 차이가 없음을 의미한다. 예를 들어 두 치료 방법의 효과를 비교하는 연구에서 검정 결과  $p=0.03$ 이라는 것은 치료효과에 차이가 없다는 귀무가설하에서 연구결과로 얻은 치료효과 차이가 우연히 나타날 확률이 3% 이하로 매우 낮기 때문에 이러한 차이는 우연에 기인한 것이 아니라 처리 즉 치료효과에 기인하였을 가능성이 높다는 것이다. 유의확률이 작을수록 귀무가설을 기각하는 증거의 강도가 높다는 것을 의미하며 이를 통계적으로 유의하다 (statistically significant)고 표현한다. 유의확률은 귀무가설을 반박하는 증거 혹은 두 군간 연관성 (association)의 강도를 측정하는 수단으로 처리효과 크기 (size of effect)에 대한 정보를 제공하는 것은 아니다.

## 7. 단측검정과 양측검정

가설검정은 양측검정 (two-tailed)과 단측검정 (one-tailed)으로 구분할 수 있다. 흔히 양측검정을 사용하지만 연구 상황에 따라서는 단측검정을 사용하는 것이 더 적절한 경우가 있다. 예를 들어 기존의 약제와 새로운 약제의 효과를 비교하는 연구에서 두 약제의 치료효과가 동일하지 않다는 차이성 검정에서는 양측검정을 사용하지만, 기존 약제에 비하여 새로운 약제의 효과가 더 크다는 것에 관심을 둔다면 단측검정을 사용해야 한다. 새로운 치료제의 치료효과가 기존의 치료제의 효과에 비하여 나쁘지 않다는 것을 검정하는 비열등성 시

험 (non-inferiority testing)에서는 단측검정을 사용하며 두 군에서 치료효과의 차이가 임상적으로 무시할 수 있는 동등성 인정한계 이내에 있는지를 검정하고 특히 새로운 치료제의 용량-반응관계를 제시하는 동등성 시험 (equivalence testing)에서는 양측검정을 사용한다. 마찬가지로 새로운 치료제가 기존의 치료제에 비하여 약제의 유효성이 더 우수하다는 것을 검정하는 우위성 검정 (superiority testing)에서는 단측검정을 사용할 수 있다. 한국임상수의학회지에 발표된 모든 논문에서는 양측검정을 사용하고 있는데 전술하였듯이 연구가설에 따라 검정의 형태가 다를 수 있기 때문에 연구목적에 가장 부합하는 가설을 세우고 분석하는 것이 바람직하다.