

충남대 학생은 학년과 성별에 따라 희망연봉에 차이가 있는가?

김지우 정철웅 김희선

충남대학교 자연과학대학 정보통계학과



동기

청년들에게는 편하고, 급여가 좋고, 장기간 일할 수 있는 직업을 곧, '좋은 직업'이라는 의식이 팽배해 있다. 따라서 공무원, 공기업, 대형 사업장만을 고집하게 되고, 실제로 장기간 실업자로 공부에만 전념하는 경우가 태반이다. 우리의 연구는 여기서 출발한다. 그렇다면 그들은 얼마만큼의 연봉을 희망하며, 학년이 올라감에 따라 의식이 달라지는가, 더불어 여성 고용이 점차 증가하는 이 시기에, 과연 남녀의 희망연봉에 차이가 있는가를 알아보기 위해, 우리는 '학년과 성별에 따라 희망연봉에 차이가 있는가?'를 주제로 조사 및 연구를 하게 되었다.

모집단설정 및 표본설계

모집단은 신분적, 시간적, 금전적 상황을 고려해 모집단 설정을 '충남대학교 재학 및 휴학생'을 대상으로 하였다.

표본설계 또한 앞서의 조건들을 고려하고, 각 단대 고유의 특성이 결과에 영향을 주는 것을 방지하기 위해 '균등배분법'에 의거, '비례증화추출법'을 선택하였다.

이에 대학본부에 문의, 각 단대별 성별 인원수 자료를 얻었다.

인문 사회 자연 경상 공과 농업생명 사범 법학 생명시스템 생활과학 약학 예술 계 총

남 1124 1903 2155 6240 1575 879 281 328 146 59 192 15901 24294

여 1625 383 1118 1284 993 686 288 112 200 484 48 562 8383

다음으로 표본크기를 결정하도록 하겠다.

95% 신뢰수준 하에 허용오차 ±5%이내로 표본크기 결정하였고, 이에

$$n_0 = z^2 \times \frac{p(1-p)}{\varepsilon^2}, n = n_0 \times \sqrt{\frac{N-n_0}{N-1}}$$

n_0 은 수정전 표본크기, z 는 신뢰수준, p 은 0.5, N 은 모집단크기, ε 는 오차범위, n 은 적정표본크기가 되겠다. 이에 계산해보면,

$$384.16 = 1.96^2 \times \frac{0.5 \times 0.5}{0.05^2}, 382 \approx 384.16 \times \sqrt{\frac{24284 - 384.16}{24284 - 1}}$$

계산결과 적정표본크기는 382가 되었으며, 우리는 최종표본크기를 400명으로 결정했다.

이에 균등배분법에 의해 성별당 한 학년 표본 수는 50명으로 설정, 비례증화추출을 위해 각 단대 주출인원을 계산해보면 다음과 같다.

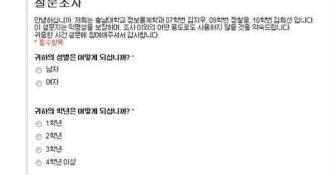
인문 사회 자연 경상 공과 농업생명 사범 법학 생명시스템 생활과학 약학 예술 계 총

남 10 6 7 19 5 3 1 1 0 1 50

여 10 6 7 6 4 2 1 1 3 0 3 50

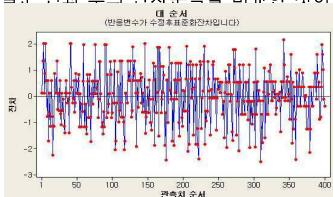
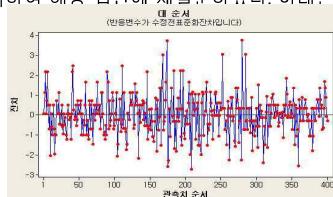
다음은 설문조사 방법을 소개하도록 하겠다.

조사방법으로 'google drive'를 이용, 인터넷 설문조사형식으로 설문하였으며, 각 단대별 학장 허가를 받고, 학생회장에게 양해를 받아 단대별로 계산된 인원만큼 랜덤추출하여 설문을 배포했다. 아래 왼쪽은 설문결과가 실시간 업데이트되는 상황이며, 오른쪽은 설문지이다.



다음은 이상치 제거 및 재설문과정이다.

이상치 판단여부는 잔차분석을 통해 판단하였고, 표준화잔차가 절대값 2.5이상인 자료를 제거하여 해당 칸에 재설문여부였다. 아래는 재설문 절차 및 재설문 결과를 비교한 것이다.



이로서 분석 가능한 형태로 표본이 추출되었으며, 이에 통계처리를 실시하고자 한다.

통계처리방법

요인, 즉 독립변수가 성별과 학년, 두 종류이므로 이원배치 분산분석을 실시한다.

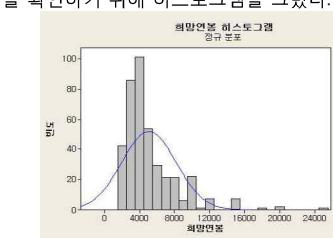
분산분석 시, 확인되어야 할 기본가정들은 정규성, 등분산성, 독립성이다.

독립성의 경우, 표본추출과정에서 이미 가정을 했기 때문에 생략하도록 하겠다.

이어 정규성 검토를 실시하도록 하겠다. 정규성 검토로 Kolmogorov-Smirnov 검정을 선택했다.

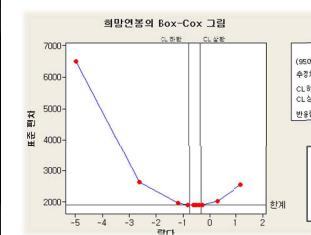
육안으로 관찰되는 것으로 볼 때, 직선에 근접하지 않은 것을 확인할 수 있다.

이에 그래프의 형태가 어떤지 대략적인 모습을 확인하기 위해 히스토그램을 그렸다.



이러한 결과가 나온 이유로 추측해본 결과, 모집단의 특성상 대상이 '희망연봉'라는 점에서 연봉은 0이하로 내려갈 수가 없기 때문에 그라프의 형태가 오른쪽으로 치우쳐진 형태가 됨을 예상할 수 있다. 즉, 원자료를 이용하여 분산분석을 실시하기에는 어려움이 따른다.

따라서 Box-Cox 변수변환을 이용하여 분석을 좀더 용이하게 변환시킬 함수를 구하고 자료들을 변수변환 시켰다.

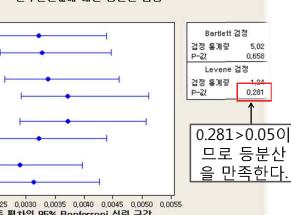


이에 변수변환된 자료들을 통해 정규성과 등분산성을 검정한다.

Calculator로 변환



변수변환값에 대한 등분산 검정



0.281>0.05이므로 등분산을 만족한다.

검정결과, 정규성과 등분산성이 만족함을 확인 할 수 있었으며, 이로서 신뢰성이 있는 분석 결과를 낼 수 있게 되었다. 그리하여 가설검정을 통해 분산분석을 실시하도록 한다.

가설검정 및 검정실시

통계모형을 $x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$ 이라 설정할 때,

x_{ijk} 는 학년의 i 번째 수준과 성별의 j 번째 수준에서 k 번째 관측된 반응값을 의미하고,

α_i 는 학년의 i 번째 효과, β_j 는 성별의 j 번째 효과,

$(\alpha\beta)_{ij}$ 는 상호작용효과라 할 수 있겠다. (ε_{ijk} 는 오차)

이에 가설검정을 세우면, 유의수준 5%하에

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$$

H_1 : 적어도 한 α_i 는 0이 아니다.

$$H_0: \beta_1 = \beta_2 = 0$$

H_1 : 적어도 한 β_i 는 0이 아니다.

$$H_0: v(\alpha\beta)_{ij} = 0$$

H_1 : 적어도 한 $(\alpha\beta)_{ij}$ 는 0이 아니다.

이에 분산분석표를 작성하면 결과는 이렇다.

이원 분산 분석: 변수변환값 대 학년, 성별

출처	DF	SS	MS	F	P
학년	3	0.0005123	0.0001708	15.37	0.000
성별	1	0.0000801	0.0000801	7.21	0.008
교학작용	3	0.0000818	0.0000273	2.45	0.063
오차	392	0.0043538	0.0000111		
총계	399	0.0050279			

분산분석 결과 해석

우선 성별과 학년의 상호작용효과(성별*학년)를 살펴보면 $P=0.063>0.05$ 이므로 통계적으로 유의하지 못하다. 즉 상호작용이 없다는 것으로 판단할 수 있다.

따라서 요인별로 주효과를 분석하고자 한다.

상호작용 효과가 없으므로 주효과의 유의확률을 보면 위의 각각의 요인의 효과가 존재함을 알 수 있는데, 이는 성별과 학년이 각각 희망연봉에 영향을 주고 있다는 것을 의미한다.

학년의 효과를 보면 $P=0.000<0.05$ 이므로,

"유의수준 5% 하에서 학년에 따른 희망연봉의 차이가 있다"고 결론을 내릴 수 있다.

마찬가지로 성별 또한 $P=0.008<0.05$ 이므로, 차이가 있다고 결론을 내릴 수 있다.

이에 사후분석을 실시한다.

사후검정 및 결과 해석

사후분석으로는 Tukey의 다중비교를 실시했다. 학년에 따른 희망연봉을 다중비교하면 결과는 이렇다.

Tukey의 방법을 사용한 그룹화 정보

학년	N	평균	그룹화	성별	N	평균	그룹화
4	100	0.016743	A	여	200	0.015902	A
3	100	0.016240	A	남	200	0.015007	B
2	100	0.015006	B				
1	100	0.013828	B				

그룹화 결과를 확인해보면, 우선 학년의 경우, 1,2학년과 3,4학년으로 비슷한 양상을 띠는 그룹끼리 나뉨을 볼 수 있다. 여기서 희망연봉의 변수변환값이 제곱근에 역수를 취한 값이기 때문에 여기서 평균값이 클수록 실제 희망연봉은 작아지는 점을 생각해 볼 때, 1,2학년에 비해 3,4학년이 좀 더 희망연봉이 낮음을 알 수 있다.

성별의 경우, 분산분석결과에서 차이가 있음을 확인할 수 있었고, 평균을 비교해보면 남자가 여자보다 조금 높음을 알 수 있다.

개인적 분석 고찰 및 마무리

학년과 성별에 따라 희망연봉의 차이가 있는가?라는 질문에 대답을 하자면 우선 답은 '그렇다'라 할 수 있겠다. 그렇다면 왜 고학년으로 갈 수록 희망연봉이 낮아지며, 여성보다 남성이 더 높은 연봉을 희망하는가 생각해봤을 때,

우선 취업을 준비하는 과정에서, 취업관련 정보를 더 많이 수집하는 쪽은 아무래도 저학년 보다 고학년일 것이다. 그에 따라 희망연봉 또한 점차 현실적인 액수로 내려가는 것이 아닌가 생각할 수 있다.

또 성별에 따라 희망연봉에 차이가 생기는 이유는, 아직도 남성이 여성보다 사회진출의 폭이 크기 때문이지 않을까 조심스레 추측할 수 있다.

물론 희망연봉의 차이로, 학과별 특성 차이나, 자신의 진로상담교수의 영향 등 수 많은 요소들이 있을 수 있다. 기회가 된다면 그러한 요소들도 고려한 분석을 해보고 싶다.