

빅데이터의 이해 HW12

- employee_data.sav의 독립표본 T 검정과 대응표본 T 검정 분석 -

201614010 영어영문학과(휴먼ICT) 박도준

employee_data.sav 파일에서 현재급여와 최초급여의 평균과, 직종에 따른 현재급여와 최초급여의 평균을 비교해보고자 한다. 두 변수의 평균을 비교할 때 사용하는 분석기법으로 T 검정이 있는데, 다른 두 집단의 평균을 비교할 시에는 독립표본 T 검정을, 집단 구분이 없는 경우에는 대응표본 T 검정을 적용하여 평균이 유의한 차이를 갖는지 비교해 볼 수 있다. 따라서 현재급여와 최초급여의 차이는 대응표본 T 검정을 적용하고, 직종에 따른 현재급여와 최초급여의 차이는 독립표본 T 검정을 적용하여 평균 값에 대한 비교를 분석해보고자 한다.

대응쌍	급여	기술통계량		T 검정		
		평균	표준 편차	t	(양쪽) 유의확률	차이의 95% 신뢰구간
1	현재급여	\$34,419.57	\$17,075.66	35.036	.000	(\$16427.41, \$18379.56)
	최초급여	\$17,016.09	\$7,870.64			

[표 1] 현재급여와 최초급여에 대한 기술통계량과 대응표본 T 검정 결과

표 1은 현재급여가 최초급여에 비해 유의하게 차이가 있는지를 비교하기 위해 대응표본 T 검정을 시행한 결과이다. 현재급여와 최초급여의 대응쌍이 유의확률 0.000으로, 두 변수의 평균은 유의한 차이가 있다고 할 수 있다. 95% 신뢰수준에서 현재급여는 최초급여에 비해 평균 급여가 작게는 \$16427에서 많게는 \$18379 정도 높음을 확인해볼 수 있다. 따라서 현재급여와 최초급여의 평균은 같다는 귀무가설을 기각하고, 두 집단의 평균이 다르다는 대립가설을 채택한다.

급여	경영직 여부	N	기술통계량		T 검정		
			평균	표준 편차	t	(양쪽) 유의확률	차이의 95% 신뢰구간
현재급여	비경영직	390	\$28,053.18	\$7,363.360	-29.396	.000	(-\$38326.00, -\$33523.24)
	경영직	84	\$63,977.80	\$18,244.780			
최초급여	비경영직	390	\$14,164.01	\$2,837.110	-27.225	.000	(-\$17255.44, -\$14932.25)
	경영직	84	\$30,257.86	\$9,980.980			

[표 2] 경영직 여부에 따른 현재급여와 최초급여의 기술통계량과 T 검정결과

표 2는 경영직 여부에 따른 현재급여와 최초급여가 유의한 차이를 보이는지 알기 위해 시행한 독립표본 T 검정의 분석 결과이다. T 검정의 유의확률을 보면 현재급여와 최초급여 모두 0.000으로 경영직 여부에 따른 유의한 급여의 차이가 존재함을 확인해 볼 수 있다. 기술통계량의 평균 급여를 보면 경영직이 비경영직보다 확연히 높음을 알 수 있다. 95%의 신뢰수준에서 현재급여는 경영직이 비경영직보다 작게는 \$33523.24에서 많게는 \$38326.00까지 높다고 할 수 있으며, 최초급여는 경영직이 비경영직보다 작게는 \$14932.25에서 많게는 \$17255.44까지 높다고 할 수 있다. 따라서 경영직 여부에 따른 현재급여와 최초급여의 평균은 같다는 귀무가설을 기각하고 두 모집단의 평균이 다르다는 대립가설을 채택한다.