



# DATA ANALYSIS REPORT

CJHSTAT CONSULT



COMMUNICATION



INTUITION



ACCURACY

## 0. 데이터

[https://www.kaggle.com/rhuebner/human-resources-data-set?select=HRDataset\\_v13.csv](https://www.kaggle.com/rhuebner/human-resources-data-set?select=HRDataset_v13.csv)

캐글에서 Human Resources 데이터를 사용함.

### 0.1. 데이터 변수 설명

변수명	설명	데이터 유형
Employee Name	직원의 full name	문자열
EmpID	직원별 고유한 id값	문자열
PayRate	시급	소수점
Sex	1 = 남성, 0 = 여성	이진
MaritalDesc	혼인 상태 (이혼 / 기혼 / 별거 / 싱글 / 과부)	문자열
PerformanceScore	성과점수 범주 (Fully Meets, Partially Meets, PIP, Exceeds) PIP는 Performance Improvement Plan을 의미함	문자열
EngagementSurvey	가장 최근 시행한 참여도 점수 설문조사	소수점
EmpSatisfaction	가장 최근 시행한 만족도 점수 설문조사	정수
SpecialProjectsCount	특별 프로젝트 수행 건수	정수

### 1. T-검정

$H_0$  : 성별에 따른 시급의 차이가 존재하지 않는다.

$H_1$  : 성별에 따라 시급의 차이가 존재한다.

Welch Two Sample t-test					95% 신뢰구간		
종속변수	유형	t 통계량	자유도	p-value	평균차이	하한	상한
PayRate	t-test	2.3771	267.84	<.05	-4.225	0.726	7.724

$H_0$ 는 귀무가설,  $H_1$ 은 대립가설을 의미함. 분석결과 유의수준 95%에서 성별에 따른 시급의 차이는 존재하는 것으로 밝혀졌으며, 평균적으로 남성이 여성보다 시간당 4.225달러 더 받는 것으로 분석 되었다.

### 2. 교차분석

$H_0$  : 혼인상태와 업무성과등급은 서로 독립이다.

$H_1$  : 혼인상태와 업무성과등급 간 연관관계가 존재한다.

교차 빈도표	PerformanceScore				
MaritalDesc	Exceeds	Fully Meets	Needs Improvement	PIP	총합
Divorced	5	23	2	0	30
Married	11	100	7	5	123
Separated	2	9	0	1	12
Single	18	104	9	6	137
Widowed	1	7	0	0	8
총합	37	243	18	12	310

카이제곱 검정			
	값	자유도	p-value
$\chi^2$	5.764	12	0.9275
N	310		

교차분석 결과,  $\chi^2$  통계량을 기준으로 유의수준 95%에서 업무성과등급과 혼인상태는 서로 독립인 것으로 밝혀

졌다. 따라서 직원들의 업무성과등급은 혼인상태와 연관성이 없다고 본다.

3. 상관관계분석

		PayRate	SpecialProjectsCount	EngagementSurvey	EmpSatisfaction
PayRate	Pearson's r	—			
	p-value	—			
SpecialProjectsCount	Pearson's r	0.4485	—		
	p-value	<.05			
EngagementSurvey	Pearson's r	-0.0242	-0.0403	—	
	p-value	0.671	0.48	—	
EmpSatisfaction	Pearson's r	0.0615	0.0365	-0.0171	—
	p-value	0.2802	0.5225	0.764	

상관분석 결과, 유의수준 95%에서 시급과 특별프로젝트건수 간 중간(0.4485) 정도의 양의 선형 상관관계가 존재한다. 그러나, 참여도 점수 및 만족도 점수와의 상관관계는 유의하지 않았다. 특별프로젝트건수, 참여도 점수, 만족도 점수 간 상관관계는 모두 유의하지 않았다.

4. 다중회귀분석

모형의 설명력

R	R²
0.4588	0.2105

모형의 유의성 검정

F 통계량	df1	df2	p
40.92	2	307	<.05

회귀분석 적합결과

독립 변수	추정B	표준편차	t 통계량	p-value
(Intercept)	26.514	1.086	24.41	< .05
SexM	3.008	1.58	1.904	0.0578
SpecialProjectsCount	2.877	0.333	8.639	<.05

4.1. 모형: 시급 ~ 성별 + 특별프로젝트건수

시급을 종속변수로, 성별 및 특별프로젝트건수를 독립변수로 다중회귀분석을 실시하였다. F검정 결과 회귀모형은 유의수준 95%에서 유의했으며, 모형으로 설명되는 데이터의 분산은 21.05%로 계산되었다. 독립변수의 효과를 추정한 결과 유의수준 95%에서 성별의 효과는 유의하지 않았지만 특별프로젝트건수의 효과는 존재했다. 특별프로젝트건수가 1건 증가하면 시급은 평균적으로 2.877달러 상승했다. 만약 유의수준을 90%로 설정한다면, 성별의 효과도 유의하며 남성의 시급이 여성보다 평균적으로 약 3달러 더 높다고 볼 수 있다.