

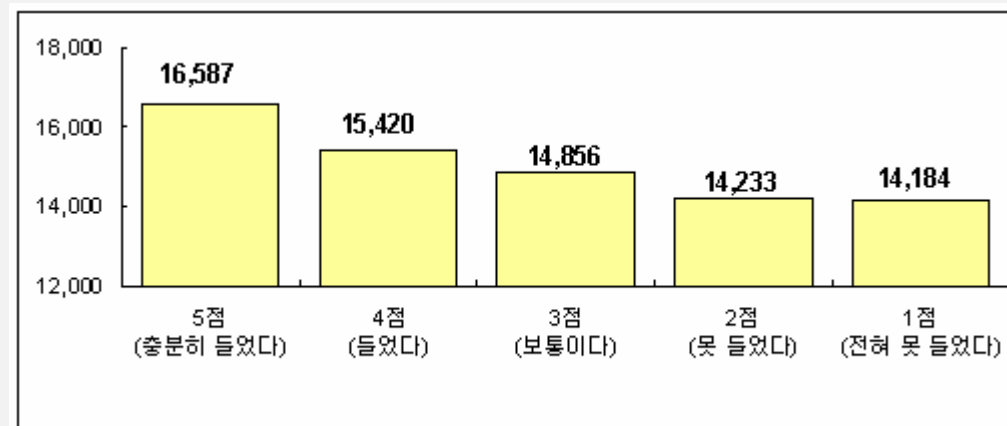


제12장 집단간 차이분석을 통한 인과관계검증

| 도입 사례 |

D화재에서는 자사의 마케팅 성과에 영향을 미치는 요인들을 파악하여 마케팅전략의 수립 및 실행에 사용해오고 있습니다. D화재에서는 자사가 판매하는 보험 상품의 기존 소비자들을 대상으로 '계약관리 만족도' 설문조사를 시행하였고 총 9,248건의 응답을 확보하였습니다. 동 설문조사 자료를 바탕으로 "상품 설명 정도에 따라 계약 건당 특약보험료의 차이가 있다"라고 하는 가설을 설정하고 이를 분산분석을 통해 분석한 내용은 아래의 그림과 같습니다.

<상품설명정도에 따른 건당 평균 특약보험료>





제12장 집단간 차이분석을 통한 인과관계검증

생각해 볼 문제_____

- ① D화재의 분석에 의하면 분산분석은 $=0.017$ 수준에서 유의하다고 합니다. 즉 5개 집단 간에 특약보험료가 동일하지는 않다는 것이지요. 그렇다면 어떠한 집단 간(예:1점 집단과 2점 집단 간)에 차이가 있는 것일까요? 또 차이가 있는지 여부는 어떻게 판단할까요?
- ② 상품설명정도가 특약보험료의 차이를 야기한다면, 이러한 정보를 마케팅 성과제고에 어떻게 이용할 수 있을까요?



1. 인과관계의 정의

- ◆ 인과관계/인과성(**causation** 혹은 causality)
 - “사물의 생성·변화에서 원인과 결과의 관계”로 사전적(辭典的)으로 정의됨
 - ☞ 과학이나 철학 분야에 있어서 많은 논쟁거리를 만들고 있는 주제
 - ☞ 우리가 다룰 인과관계는 결정론적 관계(deterministic relationship)가 아닌 확률적 관계(probabilistic/stochastic relationship)로 이해
 - ☞ 인과관계에 대한 추론(inference)만이 가능
 - **본서**에서는 “이론이 예측하는 방향과 크기만큼 독립변수의 변화가 종속변수의 변화를 야기하는 현상”으로 한정



2. 인과관계 성립조건: [1], [2], [3]

1) 두 개 현상의 동시변화 조건(concomitant variation 혹은 covariation)

- ▣ 변화하는 현상들간의 관계, 즉 변수(variable)와 변수(variable)의 관계만 인과관계 추론의 대상이 됨

☞ 변화는 현상의 본질?

예) 주역/역경(The Book of Changes)

☞ 두 개의 현상이 변한다고 함을 객관적으로 판단할 수 있기 위해 측정과정 필요

※ 상관관계분석(10장)은 두 개(이상)의 현상이 동시에 변하는가를 분석해주는 기법

2) 원인선행조건(time order of the occurrences)

- ▣ 원인변수(=독립변수)는 결과변수(=종속변수)보다 선행(先行)해야 함

예: 광고비 변동이 매출액 변동을 선행



3) 기타원인 부재조건(nonspuriousness)

- 오로지 원인(예: X)만이 결과(예: Y)를 변화시켜야 하고, 제3의 변수가 결과에 영향을 줄 가능성은 모두 제거되어야 함

예) 소방관의 수 → 화재피해? vs. 화재의 크기 → 소방관의 수/화재피해?

☞ 완벽한 확인이 불가능에 가까운 조건임(∵ 연구자가 생각지 못한 제3의 변수가 존재할 가능성이 항상 존재)

☞ 결국, 인과관계에 대한 추론은 정도의 문제(matter of degree)

-> 실험설계를 이용하여 상당수준의 인과관계 추론이 가능



3. 인과관계분석에 적합한 연구설계

◆ 실험설계(experimental design)

- 조사자가 독립변수(즉, 원인으로 생각되는 현상)를 조작(manipulate), 통제(control)하면서 종속변수(즉, 결과로 생각되는 현상)를 관찰하는 연구설계
- 독립변수의 변화만이 종속변수의 변화를 야기하는지 알아보는 연구방법

* 연구설계 : 연구문제에 대한 답을 구하도록 짜여진 연구의 계획 및 구조

-> 어떠한 현상들을, 어떻게 관찰하고, 관찰결과를 어떻게 분석할지를 계획하는 것

1) 실험설계의 기본모형

실험의 기본모형

(식 12-1)

실험집단: $Y_1 \rightarrow X \rightarrow Y_2$

통제집단: $Y_3 \rightarrow Y_4$

단, X 란 독립변수의 조작, 즉 독립변수를 실험집단에 적용함을 의미

Y_1 과 Y_3 는 실험전의 종속변수의 수준

Y_2 와 Y_4 란 실험후의 종속변수의 수준



제1절 인과관계분석이란 무엇인가?

- ◆ 실험집단(experimental group) : 연구자가 독립변수를 적용하는 집단
- ◆ 통제집단(control group) : 독립변수를 적용하지 않는 집단
- ◆ 위약효과(placebo effect)
 - ☞ 통제집단의 필요성

2) 실험설계의 기본요소

(1) 독립변수의 조작 및 종속변수의 측정

* 영어과외

- ① 실험설계가 실행되기 위해서는 연구자가 독립변수를 조작할 수 있어야 함
 - ☞ 연구자 임의로 독립변수의 변화의 강도를 조절할 수 있어야 한다는 것
 - ☞ 인과관계추론의 1조건 (두 개 현상의 동시변화 조건)(종속변수에 대한 측정이 가능하고 변화 정도가 확인되는 경우 종속변수도 변화하는 것임을 확인)
- ② 실험 전 & 후에 종속변수를 측정할 수 있어야 함
 - ☞ 실험 전 & 후 측정치에 변화가 있는 경우 인과관계추론의 1조건 충족
 - ☞ 종속변수상의 변화보다는 독립변수의 변화가 선행하게 되면 인과관계 추론의 2조건(원인선행조건)도 성립



제1절 인과관계분석이란 무엇인가?

(2) 외생변수의 통제와 실험대상의 무작위화

◆ 외생변수(extraneous variable)

* 영어과외_IQ 높은 집단

- 독립변수를 제외하고 종속변수에 영향을 미치는 모든 변수

◆ 외생변수의 영향을 최소화하는 실험설계 필요

☞ 실험대상의 무작위(randomization) 선발

☞ 완벽한 무작위화시 인과관계 추론의 3조건(기타원인 부재조건)도 성립

☞ 완벽한 무작위화는 현실이라기보다는 이상(ideal)

3) 실험설계의 유형

〈표 12-1〉 실험설계의 유형

	원시실험설계	순수실험설계	준실험설계
독립변수 조작가능성	일부 가능	가 능	일부 가능
외생변수 통제가능성	불가능	가 능	일부 가능
실험대상 무작위화	불가능	가 능	불가능

(1) 원시실험설계(pre-experimental design)

- ▣ 실험의 일부 형식을 갖추었으되(독립변수의 조작 및 종속변수의 측정), 하나의 집단만을 대상으로 실험이 이루어지거나, 실험대상집단의 무작위화가 없는 상태에서 실험이 이루어지는 실험설계 유형
 - ☞ 각종 외생변수의 위험에 노출

(2) 순수실험설계(true experimental design)

- ▣ 실험의 기본요소를 모두 갖춘 실험의 유형
 - 예) 사전사후 통제집단 설계(pretest-posttest control group design)



제1절 인과관계분석이란 무엇인가?

사전사후 통제집단 설계

(식 12-2)

실험집단(R): $Y_1 \rightarrow X \rightarrow Y_2$

통제집단(R): $Y_3 \rightarrow Y_4$

☞ $(Y_2 - Y_1) - (Y_4 - Y_3) = \text{독립변수의 효과} + \text{시험효과}(\text{testing effect})$

☞ 솔로몬의 4집단 설계(Solomon 4-group design)로 시험효과도 제거가능

(3) 준실험설계(quasi-experimental design)

- ▣ 실험이 아닌 실제 상황에서 독립변수를 조작하는 실험설계의 유형
- ▣ 현장실험설계(field experimental design)
 - ☞ 순수실험설계보다 현실성과 외적 타당성이 높음
 - ☞ 독립변수의 조작이나 외생변수의 통제가 어려워져 변수간의 인과관계 추론을 다루는 내적 타당성이 낮은 실험설계

* 설문조사

- 내적 타당성 VS. 외적 타당성

* 추리통계분석은 인과관계 추론을 위한 조건 중 단지 일부 증거만을 제시해

주는 역할을 하는 것(인과관계의 추론은 까다로운 기준을 여러 개 충족시켜야 하는 매우 어려운 작업임)

-> 통계적 분석결과에만 의존해서 독립변수와 종속변수간의 인과관계를 추론할 수는 없는 것



1. 집단간 차이분석이란?

◆ 집단간 차이분석

- 통계분석기법이 달성하려고 하는 목표를 지칭

☞ 사회현상들간의 인과관계를 추론함에 있어서 어떠한 원인이 집단간의 차이를 야기했는지를 알아볼 필요성은 많음 (어느 정도의 변화를 유도하는지에 대한 답은 주지 못함, 오직 독립변수의 효과가 있는지만 알려줄 뿐)

예) 신약의 효과, 과외의 효과, 광고의 효과 등등

☞ 인과관계 추론이 아니면서도 집단간 차이분석이 필요할 경우도 있음

예) 성별, 연령별, 소득수준별, 지역별 차이

◆ 집단간 차이분석을 수행하는 통계기법

- t -검정, z -검정, 분산분석 등

2. t -검정이란 무엇인가?

◆ t -검정(t -tests)

- 표본의 크기가 작고 표준편차가 알려지지 않은 경우에 t 분포를 사용해서 수행하는 단 일변량 가설검정(univariate hypothesis test)방법

◆ t -검정의 유형

① 독립표본 t -검정(independent sample t -test)

- 두 집단간의 차이(평균, 비율 등)를 분석해 주는 t -검정의 유형
 - 두 집단의 모집단이 정규분포를 이루고 두 집단의 분산이 동일하다는 가정 필요
-> 분산이 동일하지 않은 경우에 대처하는 수단도 개발되어 있음
 - 집단을 구분하는 변수(독립변수)는 명목척도의 수준에서 측정된 자료이고, 종속변수는 평균, 비율 등의 계산이 가능한 등간척도 이상의 척도로 측정된 자료이어야 t -검정이 의미를 가짐

② 동일표본 t -검정(paired sample t -test) = 대응표본 t -검정

- 하나의 집단에 대해 독립변수를 적용하기 전과 적용한 후의 종속변수의 수준을 측정 하고 이들의 차이가 통계적으로 의미가 있을 만큼 큰 것인가를 분석해 주는 t -검정의 유형 예) 광고 전후의 상품 선호도 측정

③ 단일표본 t -검정(one sample t -test)

- 하나의 표본평균이 예측된 평균과 같은지 여부도 분석해 주는 t -검정의 유형

예) 신상품 idea에 대한 소비자들의 선호도가 10점 만점에 7점 이상이면 이 idea를 상품화하겠다는 결정을 할 경우, 하나의 표본에서의 선호도가 7점 이상인지를 단일표본 t -검정으로 가설검정을 할 수 있을 것임



3. 분산분석이란 무엇인가?

- ◆ 분산분석(analysis of variance: ANOVA)
 - 명목척도 수준으로 측정된 독립변수가 등간척도나 비율척도 수준으로 측정된 종속변수에 미치는 영향을 분석하는 통계분석기법의 하나
 - ☞ 하나의 통계분석기법이기보다는 통계분석에 대한 접근법(approach) 혹은 사고의 방법(way of thinking)이기도 함
- ◆ t -검정과 분산분석의 차이점
 - t -검정은 두 개의 집단간에 존재하는 종속변수상의 차이를 분석해 주는 기능을 수행함에 비해, **분산분석**은 두 개 이상 n개의 집단간에 존재하는 종속변수상의 차이를 분석해 주는 기능을 수행
 - 예) 집단 1, 2, 3 사이에 존재하는 차이 분석시 독립표본 t -검정을 실시하려면 총 ${}_3C_2(=3)$ 개의 집단 비교가 필요하나, (일원)분산분석을 사용하면 단 한 번의 분산분석으로도 집단간에 차이가 존재하는지 알아낼 수 있음



◆ 분산분석의 기본원리

- 종속변수의 변화폭(즉, 분산)이 독립변수에 의해 주로 기인하는지를 밝혀내는 것
- 종속변수의 변화폭이 우연(즉, 설명 불가능한 변동폭)보다 필연(즉, 독립변수)에 의해 야기되고 이러한 변화폭이 충분히 큰지를 분석해 주는 것
- 집단간 분산(V_b)이 집단내 분산(V_w)보다 충분히 큰 것인가를 판단하는 것

분산의 구성

(식 12-3)

$$\begin{aligned} V_t = \sigma^2 &= \frac{\sum (X_i - \bar{X})^2}{N} \\ &= \text{체계적 분산}(V_s) + \text{비체계적 분산}(V_{us}) \\ &= V_s + V_{us} \\ &= \text{집단간 분산}(V_b) + \text{집단내 분산}(V_w) \\ &= V_b + V_w \\ &= \text{실험분산}(V_{exp}) + \text{오차분산}(V_e) \\ &= V_{exp} + V_e \end{aligned}$$



제2절 집단간 차이분석이란 무엇인가?

◆ 분산분석의 유형

- ① **일원분산분석**(one-way analysis of variance: one-way ANOVA)
 - 완전무작위설계(completely randomized design: CRD), 일원배치분산분석
- ② **공변량분석**(analysis of covariance: ANCOVA)
- ③ **다원분산분석**(n -way analysis of variance 혹은 multi-way analysis of variance: n -way ANOVA)
 - Factorial design, 독립변수들간의 상호작용효과도 분석 가능
- ④ **다변량분산분석**(multivariate analysis of variance: MANOVA)

〈표 12-2〉 분산분석의 유형

	독립(실험)변수 수	외생변수 수	종속변수 수
일원분산분석	1	0	1
다원분산분석	2 이상	0	1
다변량분산분석	2 이상	0	2 이상
공변량분석	1	1 이상	1



제2절 집단간 차이분석이란 무엇인가?

◆ 분산분석의 기본 가정

- ① 각 집단의 모집단은 정규분포를 따라야 함
(→ 표본 크기가 크다면, CLT(중심극한정리)가 적용되어 분산분석이 가능)
- ② 각 집단의 모집단에서의 분산은 동일(homogeneity)해야 함
(→ 각 집단의 표본 크기가 동일하면 이런 가정 위배 되어도 분산분석 적용 가능)
- ③ 각 집단은 각자의 모집단에서 무작위로 추출되고 상호 독립적인 표본이어야 함
(* 수집된 자료가 이 조건 위배시 분산분석 해석에 큰 오류가 발생)



1. 독립표본 t -검정

◆ 독립표본 t -검정

- 표본을 대상으로 관찰한 변수들의 측정값으로 표본통계량을 계산하고, 표본통계량과 해당 표본통계량의 표본분포(sampling distribution)에 근거해서 모집단에서도 예측된 관계가 성립하는지에 대한 가설을 검정

<표 12-3> 독립표본 t -검정시의 실질가설: 양측검정

연구가설(H_1): 두 집단의 평균(= 종속변수)은 동일하지 않을 것이다.

귀무가설(H_0): 두 집단의 평균(= 종속변수)은 동일할 것이다.

<표 12-4> 독립표본 t -검정시의 실질가설: 단측검정

연구가설(H_1): 집단 1의 평균이 집단 2의 평균보다 클 것이다.

귀무가설(H_0): 집단 1의 평균이 집단 2의 평균과 같거나 작을 것이다.



제3절 집단간 차이분석 결과는 어떻게 해석할 것인가?

◆ SPSS Statistics를 통해서 독립표본 t -검정을 실시한 결과(<표 12-6>)

- ☞ 분석 대상 변수: '현재급여'와 '성별'
- ☞ 연구가설(H_1): "성별에 따라 현재급여의 차이가 있을 것이다"
- ☞ 귀무가설(H_0): "성별에 따른 현재급여의 차이는 없을 것이다"

<표 12-5> 독립표본 t -검정시 통계가설

$$H_1: \mu_f \neq \mu_m$$

$$H_0: \mu_f = \mu_m$$

단, μ_f = 모집단에서의 여성의 현재급여 평균

μ_m = 모집단에서 남성의 현재급여 평균

◆ 독립표본 t -검정이 수행하는 가설검정 내용

- ☞ $H_0: \mu_f = \mu_m$ 라는 가정하에서, 표본에서의 두 집단간 평균의 차이(= $\bar{X}_f - \bar{X}_m$)가 모집단의 평균의 차이(= $\mu_f - \mu_m$)로부터 얼마나 멀리 떨어져 있는지를 판단하는 것



제3절 집단간 차이분석 결과는 어떻게 해석할 것인가?

두 집단이 동일 분산인 경우의 독립표본 t -검정의 검정통계량 (식 12-4)

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s(\bar{X}_1 - \bar{X}_2)}$$

단, n_1 = 집단 1의 표본크기

n_2 = 집단 2의 표본크기

$$s(\bar{X}_1 - \bar{X}_2) = \sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \text{검정통계량의 표준편차}$$

$$s^2 = \frac{\sum_{i=1}^{n_1} (X_{i1} - \bar{X}_1)^2 + \sum_{i=1}^{n_2} (X_{i2} - \bar{X}_2)^2}{n_1 + n_2 - 2} = \text{두 개의 집단을 합한 전체표본}(=n_1+n_2)\text{의 분산}$$

두 집단의 분산이 다를 경우 독립표본 t -검정의 검정통계량 (식 12-5)

$$s(\bar{X}_1 - \bar{X}_2) = \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)}$$

검정통계량 t 의 근사값을 계산하게 됨, 이 경우의 자유도는 정수가 아닐 가능성이 큼

제3절 집단간 차이분석 결과는 어떻게 해석할 것인가?

<표 12-6> 독립표본 t -검정결과

집단통계량

	성별	N	평균	표준편차	평균의 표준오차
현재급여	남자	258	\$41,441.78	\$19,499.214	\$1,213.968
	여자	216	\$26,031.92	\$7,558.021	\$514.258

독립표본 검정

		Levene의 등분산 검정		평균의 동일성에 대한 t -검정						
		F	유의확률	t	자유도	유의확률 (양쪽)	평균차	차이의 표준오차	차이의 95% 신뢰구간	
									하한	상한
현재급여	등분산이 가정됨	119.669	.000	10.945	472	.000	\$15,409.862	\$1,407.906	\$12,643.322	\$18,176.401
	등분산이 가정되지 않음			11.688	344.262	.000	\$15,409.862	\$1,318.400	\$12,816.728	\$18,002.996



제3절 집단간 차이분석 결과는 어떻게 해석할 것인가?

◆ 결과 해석시 두 집단의 분산이 동일한지 여부확인 필요

- ☞ 'Levene의 등분산 검정'이라는 방법을 이용
- ☞ 등분산의 검정에 사용된 F 값 119.669에 대응하는 유의확률(= p 값)이 0.000
- ☞ $\alpha=0.01$ 혹은 $\alpha=0.05$ 수준에서 두 집단은 동일 분산이라는 귀무가설($H_0: \sigma_1^2 = \sigma_2^2$)이 기각
- ☞ 등분산이 가정되지 않은 경우의 t 검정통계량 값(=11.688)과 해당 유의확률($p = 0.000$)을 해석

◆ 등분산이 가정되지 않은 경우의 t 검정통계량에 대응하는 유의확률 = 0.000

- ☞ $\alpha = 0.01$ 수준에서 $H_0: \mu_f = \mu_m$ 라는 귀무가설을 안전하게 기각
- ☞ 성별이 다르면 현재급여가 달라진다는 결론



2. 일원분산분석

◆ 분산분석

- ▣ 표본을 대상으로 관찰한 변수들의 측정값으로 표본통계량을 계산하고, 표본통계량과 해당 표본통계량의 표본분포(sampling distribution)에 근거해서 모집단에서도 관계가 성립하는지에 대한 가설을 검정

<표 12-7> 일원분산분석(one-way ANOVA) 시의 실질가설: 양측검정

연구가설(H_1): c 개 집단의 평균(=종속변수)이 동일하지는 않을 것이다.

귀무가설(H_0): c 개 집단의 평균(=종속변수)은 동일할 것이다.

◆ 일원분산분석: 가장 기본적인 분산분석

= 일원배치분산분석(SPSS Statistics 내에서의 표현)

◆ SPSS Statistics를 통해서 일원분산분석을 실시한 결과(<표 12-8>)

- ☞ 분석 대상 변수 : '직종구분'과 '현재급여'
- ☞ 연구가설(H_1) : "직종구분에 따라 현재급여의 차이가 있을 것이다"
- ☞ 귀무가설(H_0) : "직종구분에 따른 현재급여의 차이는 없을 것이다"



제3절 집단간 차이분석 결과는 어떻게 해석할 것인가?

<표 12-9> 일원분산분석(one-way ANOVA)시의 통계가설

$$H_0: \mu_c = \mu_a = \mu_m$$

단, μ_c = 모집단에서의 사무직(clerk)의 현재급여 평균

μ_a = 모집단에서 관리직(administrator)의 현재급여 평균

μ_m = 모집단에서 경영직(management)의 현재급여 평균

◆ 일원분산분석이 수행하는 가설검정 내용

☞ $H_0: \mu_c = \mu_a = \mu_m$ 라는 가정하에서, 모집단에서의 집단간 분산과 집단내 분산의 비율(V_b/V_w)이 충분히 큰지를 판단하는 것

일원분산분석의 검정통계량

(식 12-6)

F = 모집단의 집단간분산 추정값 / 모집단의 집단내분산 추정값

$$= S_b^2 / S_w^2 = \frac{SSB / (c - 1)}{SSW / (N - c)} = \frac{MSB}{MSW}$$

$$= \frac{\sum_{j=1}^c N(\bar{Y}_j - \bar{Y})^2 / (c - 1)}{\sum_{j=1}^c \sum_{i=1}^N (Y_{ij} - \bar{Y}_j)^2 / (N - c)}$$

단, \bar{Y}_j = 집단 j 의 종속변수 평균

\bar{Y} = 모든 집단을 더한 전체 집단에서의 종속변수 평균

Y_{ij} = j 집단에 속하는 i 응답자의 종속변수 값

c = 독립변수를 구성하는 집단의 수

N = 전체 집단의 크기



제3절 집단간 차이분석 결과는 어떻게 해석할 것인가?

S_b^2 = 집단간 자승합(between groups sum of squares: SSB)
= 모집단의 집단간 분산에 대한 추정값
= 집단의 종속변수 평균이 전체평균에서 떨어진 정도(집단간 변량[between groups variation])

S_w^2 = 집단 내 자승합(within group sum of squares: SSW)
= 모집단의 집단 내 분산에 대한 추정값
= 각 집단에 속하는 각 응답자의 종속변수값이 각 집단의 평균에서 떨어진 정도(집단 내 변량[within group variation])

MSB = 집단간 자승평균(between groups mean squares)
= 집단의 수를 고려했을 때 독립변수가 종속변수에 미치는 영향을 나타내는 '집단간 평균변량'

MSW = 집단 내 자승평균(within group mean squares)
= 표본의 크기를 고려했을 때 각 집단 내에서 독립변수와 무관하게 발생하는 변화라고 할 수 있는 '집단 내 평균변량'

☞ 모집단의 분산의 구성비(=집단간분산/집단내분산)가 충분히 큰지를 판단하기 위하여 분산개념과 유사한 집단간 변량과 집단내 변량을 표본에서 계산한 다음 <식 12-6>에서처럼 F값을 추정하게 되는 것임



<표 12-8> 일원배치분산 분석결과

기술통계

현재급여

	N	평균	표준편차	표준오차	평균에 대한 95% 신뢰구간		최소값	최대값
					하한값	상한값		
사무직	363	\$27,838.54	\$7,567.995	\$397.217	\$27,057.40	\$28,619.68	\$15,750	\$80,000
관리직	27	\$30,938.89	\$2,114.616	\$406.958	\$30,102.37	\$31,775.40	\$24,300	\$35,250
경영직	84	\$63,977.80	\$18,244.776	\$1,990.668	\$60,018.44	\$67,937.16	\$34,410	\$135,000
합계	474	\$34,419.57	\$17,075.661	\$784.311	\$32,878.40	\$35,960.73	\$15,750	\$135,000

분산의 동질성에 대한 검정

현재급여

Levene 통계량	자유도1	자유도2	유의확률
59.733	2	471	.000

분산분석

현재급여

	제곱합	자유도	평균제곱	F	유의확률
집단-간	89438483925.943	2	44719241962.971	434.481	.000
집단-내	48478011510.397	471	102925714.459		
합계	137916495436.340	473			



사후검정

다중 비교

종속변수: 현재급여

	(I) 직종구분	(J) 직종구분	평균차 (I-J)	표준오차	유의확률	95% 신뢰구간	
						하한값	상한값
Tukey HSD	사무직	관리직	-\$3,100,349	\$2,023,760	,277	-\$7,858,50	\$1,657,80
		경영직	-\$36,139,258(*)	\$1,228,352	,000	-\$39,027,29	-\$33,251,22
		관리직	\$3,100,349	\$2,023,760	,277	-\$1,657,80	\$7,858,50
	경영직	사무직	-\$33,038,909(*)	\$2,244,409	,000	-\$38,315,84	-\$27,761,98
		사무직	\$36,139,258(*)	\$1,228,352	,000	\$33,251,22	\$39,027,29
		관리직	\$33,038,909(*)	\$2,244,409	,000	\$27,761,98	\$38,315,84
Scheffe	사무직	관리직	-\$3,100,349	\$2,023,760	,310	-\$8,069,80	\$1,869,10
		경영직	-\$36,139,258(*)	\$1,228,352	,000	-\$39,155,54	-\$33,122,98
		관리직	\$3,100,349	\$2,023,760	,310	-\$1,869,10	\$8,069,80
	경영직	사무직	-\$33,038,909(*)	\$2,244,409	,000	-\$38,550,17	-\$27,527,65
		사무직	\$36,139,258(*)	\$1,228,352	,000	\$33,122,98	\$39,155,54
		관리직	\$33,038,909(*)	\$2,244,409	,000	\$27,527,65	\$38,550,17
Dunnett T3	사무직	관리직	-\$3,100,349(*)	\$568,679	,000	-\$4,482,16	-\$1,718,53
		경영직	-\$36,139,258(*)	\$2,029,912	,000	-\$41,074,92	-\$31,203,59
		관리직	\$3,100,349(*)	\$568,679	,000	\$1,718,53	\$4,482,16
	경영직	사무직	-\$33,038,909(*)	\$2,031,840	,000	-\$37,979,39	-\$28,098,43
		사무직	\$36,139,258(*)	\$2,029,912	,000	\$31,203,59	\$41,074,92
		관리직	\$33,038,909(*)	\$2,031,840	,000	\$28,098,43	\$37,979,39

* .05 수준에서 평균차가 큼니다.



제3절 집단간 차이분석 결과는 어떻게 해석할 것인가?

동일 집단군
현재급여

	직종구분	N	유의수준 = .05에 대한 부집단	
			1	2
Tukey HSD(a,b)	사무직	363	\$27,838.54	
	관리직	27	\$30,938.89	
	경영직	84		\$63,977.80
	유의확률		.227	1.000
Scheffe(a,b)	사무직	363	\$27,838.54	
	관리직	27	\$30,938.89	
	경영직	84		\$63,977.80
	유의확률		.259	1.000

동일 집단군에 있는 집단에 대한 평균이 표시됩니다.

a: 조화평균 표본 크기 = 58.031을(를) 사용

b: 집단 크기가 같지 않습니다. 집단크기의 조화평균이 사용됩니다. 제1종 오류 수준은 보장할 수 없습니다.



제3절 집단간 차이분석 결과는 어떻게 해석할 것인가?

- ◆ 결과 해석시 두 집단의 분산이 동일한지 여부확인 필요
 - ☞ 'Levene의 등분산 검정'이라는 방법을 이용
 - ☞ 등분산의 검정에 사용된 F 값 59.733에 대응하는 유의확률(= p 값)이 0.000
 - ☞ $\alpha=0.01$ 수준에서 3집단은 동일 분산이라는 귀무가설($H_0: \sigma_1^2 = \sigma_2^2 = \sigma^2$)이 기각
 - ☞ 예에서의 일원분산분석의 결과는 사실 자신 있게 받아들이기에는 문제가 있음
(분산분석의 2번째 가정을 위반함)
- ◆ 그러나 집단간 등분산 가정이 성립했다고 가정하고 분산분석결과를 해석
 - ☞ 유의확률 0.000
 - ☞ $\alpha=0.01$ 수준에서 $H_0: \mu_c = \mu_a = \mu_m$ 라는 귀무가설을 안전하게 기각
 - ☞ 직종구분이 다르면 현재급여가 달라진다는 결론
- ◆ 사후검정(post hoc comparison)
 - 집단간에 전반적인 차이가 존재한다는 결론이 나왔을 때, 구체적으로 어떤 특정한 집단간에 차이가 존재하는지 분석해주는 기법
 - 예) Scheffe, Tukey, Dunnett T3 등



1. 독립표본 t -검정의 실행

- ◆ [그림 12-1]처럼 분석대상 자료파일을 열고,



제4절 집단간 차이분석은 어떻게 실행하는가?

[그림 12-1] 분석대상 자료파일 열기

Employee data.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)

표시: 10 / 10 변수

	번호	성별	생년월일	피교육	직종	현재급여	최초급여	근무월수	경력	소수민족	변수	변수	변수	변수	변수
1	1 m		02/03/1952	15	3	\$57,000	\$27,000	98	144	0					
2	2 m		05/23/1958	16	1	\$40,200	\$18,750	98	36	0					
3	3 f		07/26/1929	12	1	\$21,450	\$12,000	98	381	0					
4	4 f		04/15/1947	8	1	\$21,900	\$13,200	98	190	0					
5	5 m		02/09/1955	15	1	\$45,000	\$21,000	98	138	0					
6	6 m		08/22/1958	15	1	\$32,100	\$13,500	98	67	0					
7	7 m		04/26/1956	15	1	\$36,000	\$18,750	98	114	0					
8	8 f		05/06/1966	12	1	\$21,900	\$9,750	98	0	0					
9	9 f		01/23/1946	15	1	\$27,900	\$12,750	98	115	0					
10	10 f		02/13/1946	12	1	\$24,000	\$13,500	98	244	0					
11	11 f		02/07/1950	16	1	\$30,300	\$16,500	98	143	0					
12	12 m		01/11/1966	8	1	\$28,350	\$12,000	98	26	1					
13	13 m		07/17/1960	15	1	\$27,750	\$14,250	98	34	1					
14	14 f		02/26/1949	15	1	\$35,100	\$16,800	98	137	1					
15	15 m		08/29/1962	12	1	\$27,300	\$13,500	97	66	0					
16	16 m		11/17/1964	12	1	\$40,800	\$15,000	97	24	0					
17	17 m		07/18/1962	15	1	\$46,000	\$14,250	97	48	0					
18	18 m		03/20/1956	16	3	\$103,750	\$27,510	97	70	0					
19	19 m		08/19/1962	12	1	\$42,300	\$14,250	97	103	0					
20	20 f		01/23/1940	12	1	\$26,250	\$11,550	97	48	0					

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료

Unicode:ON



제4절 집단간 차이분석은 어떻게 실행하는가?



[그림 12-2] 독립표본 t-검정의 실행

Employee data.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)

표시: 10 / 10 변수

번호	성별	생년월일	월수	경력	소수민족	변수	변수	변수	변수	변수
1	1 m	02/03/19	98	144	0					
2	2 m	05/23/19	98	36	0					
3	3 f	07/26/19	98	381	0					
4	4 f	04/15/19	98	190	0					
5	5 m	02/09/19	98	138	0					
6	6 m	08/22/19	98	67	0					
7	7 m	04/26/19	98	114	0					
8	8 f	05/06/19	\$21,900	\$9,750	98	0	0			
9	9 f	01/23/19	\$27,900	\$12,750	98	115	0			
10	10 f	02/13/19	\$24,000	\$13,500	98	244	0			
11	11 f	02/07/19	\$30,300	\$16,500	98	143	0			
12	12 m	01/11/19	\$28,350	\$12,000	98	26	1			
13	13 m	07/17/19	\$27,750	\$14,250	98	34	1			
14	14 f	02/26/19	\$35,100	\$16,800	98	137	1			
15	15 m	08/29/19	\$27,300	\$13,500	97	66	0			
16	16 m	11/17/19	\$40,800	\$15,000	97	24	0			
17	17 m	07/18/19	\$46,000	\$14,250	97	48	0			
18	18 m	03/20/19	\$103,750	\$27,510	97	70	0			
19	19 m	08/19/19	\$42,300	\$14,250	97	103	0			
20	20 f	01/23/19	\$26,250	\$11,550	97	48	0			

독립표본 T 검정...

IBM SPSS Statistics 프로세서 준비 완료

Unicode:ON



제4절 집단간 차이분석은 어떻게 실행하는가?

[그림 12-3] 독립표본 t-검정의 실행 창 1

Employee data.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)

표시: 10 / 10 변수

	번호	성별	생년월일	피교육	직종	현재급여	최초급여	근무월수	경력	소수민족	변수	변수	변수	변수	변수
1	1	m	02/03/1952	15	3	\$57,000	\$27,000	98	144	0					
2						8,750		98	36	0					
3						2,000		98	381	0					
4						3,200		98	190	0					
5						1,000		98	138	0					
6						3,500		98	67	0					
7						8,750		98	114	0					
8						9,750		98	0	0					
9						2,750		98	115	0					
10						3,500		98	244	0					
11						6,500		98	143	0					
12						2,000		98	26	1					
13						4,250		98	34	1					
14						6,800		98	137	1					
15	15	m	06/29/1962	12	1	\$27,300	\$13,500	97	66	0					
16	16	m	11/17/1964	12	1	\$40,800	\$15,000	97	24	0					
17	17	m	07/18/1962	15	1	\$46,000	\$14,250	97	48	0					
18	18	m	03/20/1956	16	3	\$103,750	\$27,510	97	70	0					
19	19	m	08/19/1962	12	1	\$42,300	\$14,250	97	103	0					
20	20	f	01/23/1940	12	1	\$26,250	\$11,550	97	48	0					

독립표본 T 검정

검정 변수(T): 현재급여 [현재급여]

집단 변수(G): 성별 (??)

집단 정의(D):

확인 불어넣기(P) 재설정(R) 취소 도움말

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료 Unicode:ON



[그림 12-4] 독립표본 t -검정의 실행에서 집단의 정의

—  

	번호	성별	생년월일	피교육	직종	현재급여	최초급여	근무월수	경력	소수민족	변수	변수	변수	변수	변수
1	1 m		02/03/1952	15	3	\$57,000	\$27,000	98	144	0					
2							8,750	98	36	0					
3							2,000	98	381	0					
4							3,200	98	190	0					
5							1,000	98	138	0					
6							3,500	98	67	0					
7							8,750	98	114	0					
8							9,750	98	0	0					
9										0					
10	1									0					
11	1									0					
12	1									0					
13	1									1					
14	1									1					
15	15 m		06/29/1962	12	1	\$27,500	\$13,500	97	66	0					
16	16 m		11/17/1964	12	1	\$40,800	\$15,000	97	24	0					
17	17 m		07/18/1962	15	1	\$46,000	\$14,250	97	48	0					
18	18 m		03/20/1956	16	3	\$103,750	\$27,510	97	70	0					
19	19 m		08/19/1962	12	1	\$42,300	\$14,250	97	103	0					
20	20 f		01/23/1940	12	1	\$26,250	\$11,550	97	48	0					



제4절 집단간 차이분석은 어떻게 실행하는가?



[그림 12-5] 독립표본 t-검정의 실행 창 2

Employee data.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)

표시: 10 / 10 변수

번호	성별	생년월일	피교육	직종	현재급여	최초급여	근무월수	경력	소수민족	변수	변수	변수	변수	변수
1	m	02/03/1952	15	3	\$57,000	\$27,000	98	144	0					
2					8,750		98	36	0					
3					2,000		98	381	0					
4					3,200		98	190	0					
5					1,000		98	138	0					
6					3,500		98	67	0					
7					8,750		98	114	0					
8					9,750		98	0	0					
9					2,750		98	115	0					
10					3,500		98	244	0					
11					6,500		98	143	0					
12					2,000		98	26	1					
13					4,250		98	34	1					
14					6,800		98	137	1					
15	m	06/29/1962	12	1	\$27,300	\$13,500	97	66	0					
16	m	11/17/1964	12	1	\$40,800	\$15,000	97	24	0					
17	m	07/18/1962	15	1	\$46,000	\$14,250	97	48	0					
18	m	03/20/1956	16	3	\$103,750	\$27,510	97	70	0					
19	m	08/19/1962	12	1	\$42,300	\$14,250	97	103	0					
20	f	01/23/1940	12	1	\$26,250	\$11,550	97	48	0					

독립표본 T 검정

검정 변수(T): 현재급여 [현재급여]

집단 변수(G): 성별 ('m' 'f')

집단 정의(D):

확인 불여넣기(P) 재설정(R) 취소 도움말

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료 Unicode:ON



제4절 집단간 차이분석은 어떻게 실행하는가?



[그림 12-6] 독립표본 t-검정결과 창

*출력결과1 [문서1] - IBM SPSS Statistics Viewer

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 삽입(I) 형식(O) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)

출력결과
T 검정
제목
노트
집단통계량
독립표본검정

T 검정

집단통계량

성별	N	평균	표준편차	평균의 표준오차	
현재근여	남자	258	\$41,441.78	\$19,499.214	\$1,213.968
현재근여	여자	216	\$26,031.92	\$7,558.021	\$514.258

독립표본 검정

Levene의 등분산 검정				평균의 동일성에 대한 T 검정						
		F	유의확률	t	자유도	유의확률 (양측)	평균차이	차이의 표준오차	차이의 95% 신뢰구간 하한	상한
현재근여	등분산을 가정함	119.669	.000	10.945	472	.000	\$15,409.862	\$1,407.906	\$12,643.322	\$18,176.401
현재근여	등분산을 가정하지 않음			11.688	344.262	.000	\$15,409.862	\$1,318.400	\$12,816.728	\$18,002.996



제4절 집단간 차이분석은 어떻게 실행하는가?

[그림 12-7] 분석대상 자료파일 열기

Employee data.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)

표시: 10 / 10 변수

	번호	성별	생년월일	피교육	직종	현재급여	최초급여	근무월수	경력	소수민족	변수	변수	변수	변수	변수
1	1 m	02/03/1952	15	3	\$57,000	\$27,000	98	144	0						
2	2 m	05/23/1958	16	1	\$40,200	\$18,750	98	36	0						
3	3 f	07/26/1929	12	1	\$21,450	\$12,000	98	381	0						
4	4 f	04/15/1947	8	1	\$21,900	\$13,200	98	190	0						
5	5 m	02/09/1955	15	1	\$45,000	\$21,000	98	138	0						
6	6 m	08/22/1958	15	1	\$32,100	\$13,500	98	67	0						
7	7 m	04/26/1956	15	1	\$36,000	\$18,750	98	114	0						
8	8 f	05/06/1966	12	1	\$21,900	\$9,750	98	0	0						
9	9 f	01/23/1946	15	1	\$27,900	\$12,750	98	115	0						
10	10 f	02/13/1946	12	1	\$24,000	\$13,500	98	244	0						
11	11 f	02/07/1950	16	1	\$30,300	\$16,500	98	143	0						
12	12 m	01/11/1966	8	1	\$28,350	\$12,000	98	26	1						
13	13 m	07/17/1960	15	1	\$27,750	\$14,250	98	34	1						
14	14 f	02/26/1949	15	1	\$35,100	\$16,800	98	137	1						
15	15 m	08/29/1962	12	1	\$27,300	\$13,500	97	66	0						
16	16 m	11/17/1964	12	1	\$40,800	\$15,000	97	24	0						
17	17 m	07/18/1962	15	1	\$46,000	\$14,250	97	48	0						
18	18 m	03/20/1956	16	3	\$103,750	\$27,510	97	70	0						
19	19 m	08/19/1962	12	1	\$42,300	\$14,250	97	103	0						
20	20 f	01/23/1940	12	1	\$26,250	\$11,550	97	48	0						

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료

Unicode:ON



제4절 집단간 차이분석은 어떻게 실행하는가?



[그림 12-8] 일원분산분석의 실행

Employee data.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)

보고서(P) 기술통계량(E) 표(B) 평균 비교(M) 일반선형모형(G) 일반화 선형 모형(Z) 혼합 모형(X) 상관분석(C) 회귀분석(R) 로그선형분석(Q) 신경망(W) 분류분석(E) 차원 축소(D) 척도분석(A) 비모수검정(N) 시계열 분석(T) 생존분석(S) 다중반응(U) 결측값 분석(Y) 다중대체(T) 복합 표본(L) 시뮬레이션(I) 품질관리(Q) ROC 곡선(V) 공간과 시간 모형화(S)

표시: 10 / 10 변수

번호	성별	생년월일	월수	경력	소수민족	변수	변수	변수	변수	변수
1	m	02/03/19	98	144	0					
2	m	05/23/19	98	36	0					
3	f	07/26/19	98	381	0					
4	f	04/15/19	98	190	0					
5	m	02/09/19	98	138	0					
6	m	08/22/19	98	67	0					
7	m	04/26/19	98	114	0					
8	f	05/06/19	\$21,900	\$9,750	98	0				
9	f	01/23/19	\$27,900	\$12,750	98	115	0			
10	f	02/13/19	\$24,000	\$13,500	98	244	0			
11	f	02/07/19	\$30,300	\$16,500	98	143	0			
12	m	01/11/19	\$28,350	\$12,000	98	26	1			
13	m	07/17/19	\$27,750	\$14,250	98	34	1			
14	f	02/26/19	\$35,100	\$16,800	98	137	1			
15	m	08/29/19	\$27,300	\$13,500	97	66	0			
16	m	11/17/19	\$40,800	\$15,000	97	24	0			
17	m	07/18/19	\$46,000	\$14,250	97	48	0			
18	m	03/20/19	\$103,750	\$27,510	97	70	0			
19	m	08/19/19	\$42,300	\$14,250	97	103	0			
20	f	01/23/19	\$26,250	\$11,550	97	48	0			

데이터 보기 변수 보기

일원배치 분산분석(O)...

IBM SPSS Statistics 프로세서 준비 완료

Unicode:ON



제4절 집단간 차이분석은 어떻게 실행하는가?



[그림 12-9] 일원분산분석 창 1

Employee data.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)



표시: 10 / 10 변수

	번호	성별	생년월일	피교육	직종	현재급여	최초급여	근무월수	경력	소수민족	변수	변수	변수	변수	변수
1	1 m		02/03/1952	15	3	\$57,000	\$27,000	98	144	0					
2						\$18,750	\$18,750	98	36	0					
3						\$12,000	\$12,000	98	381	0					
4						\$13,200	\$13,200	98	190	0					
5						\$21,000	\$21,000	98	138	0					
6						\$13,500	\$13,500	98	67	0					
7						\$18,750	\$18,750	98	114	0					
8						\$9,750	\$9,750	98	0	0					
9						\$12,750	\$12,750	98	115	0					
10						\$13,500	\$13,500	98	244	0					
11						\$16,500	\$16,500	98	143	0					
12						\$12,000	\$12,000	98	26	1					
13						\$14,250	\$14,250	98	34	1					
14	14 f		02/26/1949	15	1	\$35,100	\$16,800	98	137	1					
15	15 m		08/29/1962	12	1	\$27,300	\$13,500	97	66	0					
16	16 m		11/17/1964	12	1	\$40,800	\$15,000	97	24	0					
17	17 m		07/18/1962	15	1	\$46,000	\$14,250	97	48	0					
18	18 m		03/20/1956	16	3	\$103,750	\$27,510	97	70	0					
19	19 m		08/19/1962	12	1	\$42,300	\$14,250	97	103	0					
20	20 f		01/23/1940	12	1	\$26,250	\$11,550	97	48	0					

일원배치 분산분석

종속변수(E):

사원코드 [번호]

생년월일 [생년월일]

피교육년수 [피교육]

직종구분 [직종]

현재급여 [현재급여]

최초급여 [최초급여]

현 근무월수 [근무...]

입사전 타 근무월...

요인(F):

확인 불어넣기(P) 재설정(R) 취소 도움말

대비(N)... 사후분석(H)... 옵션(O)... 붓스트랩(B)...

데이터 보기 변수 보기



제4절 집단간 차이분석은 어떻게 실행하는가?



[그림 12-10] 일원분산분석시 변수의 선택

Employee data.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)

표시: 10 / 10 변수

번호	성별	생년월일	피교육	직종	현재급여	최초급여	근무월수	경력	소수민족	변수	변수	변수	변수	변수
1	1 m	02/03/1952	15	3	\$57,000	\$27,000	98	144	0					
2					\$18,750	\$18,750	98	36	0					
3					\$12,000	\$12,000	98	381	0					
4					\$13,200	\$13,200	98	190	0					
5					\$21,000	\$21,000	98	138	0					
6					\$13,500	\$13,500	98	67	0					
7					\$18,750	\$18,750	98	114	0					
8					\$9,750	\$9,750	98	0	0					
9					\$12,750	\$12,750	98	115	0					
10					\$13,500	\$13,500	98	244	0					
11					\$16,500	\$16,500	98	143	0					
12					\$12,000	\$12,000	98	26	1					
13					\$14,250	\$14,250	98	34	1					
14	14 f	02/26/1949	15	1	\$35,100	\$16,800	98	137	1					
15	15 m	08/29/1962	12	1	\$27,300	\$13,500	97	66	0					
16	16 m	11/17/1964	12	1	\$40,800	\$15,000	97	24	0					
17	17 m	07/18/1962	15	1	\$46,000	\$14,250	97	48	0					
18	18 m	03/20/1956	16	3	\$103,750	\$27,510	97	70	0					
19	19 m	08/19/1962	12	1	\$42,300	\$14,250	97	103	0					
20	20 f	01/23/1940	12	1	\$26,250	\$11,550	97	48	0					

일원배치 분산분석

종속변수(E):

현재급여 [현재급여]

요인(F):

직종구분 [직종]

확인 불여넣기(P) 재설정(R) 취소 도움말

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료 Unicode:ON



제4절 집단간 차이분석은 어떻게 실행하는가?

[그림 12-11] 일원분산분석시 사후분석 창

Employee data.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)

표시: 10 / 10 변수

번호	성별	생년월일	피교육	직종	현재급여	최초급여	근무월수	경력	소수민족	변수	변수	변수	변수	변수
1	1 m	02/03/1952	15	3	\$57,000	\$27,000	08	144	0					
2														
3														
4														
5														
6														
7														
8														
9														
10														
11														
12														
13														
14	14 f	02/26/1949	15	1	\$35,100	\$								
15	15 m	08/29/1962	12	1	\$27,300	\$								
16	16 m	11/17/1964	12	1	\$40,800	\$								
17	17 m	07/18/1962	15	1	\$46,000	\$14,250	97	48	0					
18	18 m	03/20/1956	16	3	\$103,750	\$27,510	97	70	0					
19	19 m	08/19/1962	12	1	\$42,300	\$14,250	97	103	0					
20	20 f	01/23/1940	12	1	\$26,250	\$11,550	97	48	0					

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료 Unicode:ON

일원배치 분산분석

종속변수(E): 현재급여 [현재급여]

요인(F): 직종구분 [직종]

확인 불어넣기(P) 재설정(R) 취소 도움말

일원배치 분산분석: 사후분석 - 다중비교

등분산을 가정할

☒ LSD ☐ S-N-K ☐ Waller-Duncan

☐ Bonferroni ☐ Tukey 방법 제1종/제2종 오류 비율: 100

☐ Sidak ☐ Tukey의 b ☐ Dunnett

☐ Scheffe ☐ Duncan 대조 범주(Y): 마지막

☐ R-E-G-W의 F ☐ Hochberg의 GT2

☐ R-E-G-W의 Q ☐ Gabriel

검정: ☒ 양측검정(2) ☐ < 대조(O) ☐ > 대조(N)

등분산을 가정하지 않음

☐ Tamhane의 T2 ☐ Dunnett의 T3 ☐ Games-Howell ☐ Dunnett의 C

유의수준(E): 0.05

계속(C) 취소 도움말



제4절 집단간 차이분석은 어떻게 실행하는가?

[그림 12-12] 일원분산분석시 사후분석 창에서 다중비교방법 선정

Employee data.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)

표시: 10 / 10 변수

	번호	성별	생년월일	피교육	직종	현재급여	최초급여	근무월수	경력	소수민족	변수	변수	변수	변수	변수
1	1 m		02/03/1952	15	3	\$57,000	\$27,000	00	111	0					
2															
3															
4															
5															
6															
7															
8															
9															
10															
11															
12															
13															
14	14 f		02/26/1949	15	1	\$35,100	\$								
15	15 m		08/29/1962	12	1	\$27,300	\$								
16	16 m		11/17/1964	12	1	\$40,800	\$								
17	17 m		07/18/1962	15	1	\$46,000	\$14,250	97	48	0					
18	18 m		03/20/1956	16	3	\$103,750	\$27,510	97	70	0					
19	19 m		08/19/1962	12	1	\$42,300	\$14,250	97	103	0					
20	20 f		01/23/1940	12	1	\$26,250	\$11,550	97	48	0					

일원배치 분산분석

종속변수(E): 현재급여 [현재급여]

요인(F): 직종구분 [직종]

확인 불여넣기(P) 재설정(R) 취소 도움말

일원배치 분산분석: 사후분석 - 다중비교

등분산을 가정함

☐ LSD ☐ SN-K ☐ Waller-Duncan

☐ Bonferroni ☒ Tukey 방법 제1종/제2종 오류 비율: 100

☐ Sidak ☐ Tukey의 b ☐ Dunnett

☒ Scheffe ☐ Duncan 대조 범주(Y): 마지막

☐ R-E-G-W의 F ☐ Hochberg의 GT2

☐ R-E-G-W의 Q ☐ Gabriel

검정: ☒ 양측검정(2) ☐ < 대조(Q) ☐ > 대조(N)

등분산을 가정하지 않음

☐ Tamhane의 T2 ☒ Dunnett의 T3 ☐ Games-Howell ☐ Dunnett의 C

유의수준(E): 0.05

계속(C) 취소 도움말

데이터 보기 변수 보기

IBM SPSS Statistics 프로세서 준비 완료 Unicode:ON



제4절 집단간 차이분석은 어떻게 실행하는가?

[그림 12-13] 일원분산분석 창에서 옵션선택

Employee data.sav [데이터세트1] - IBM SPSS Statistics Data Editor

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)

표시: 10 / 10 변수

번호	성별	생년월일	피교육	직종	현재급여	최초급여	근무월수	경력	소수민족	변수	변수	변수	변수	변수
1	1 m	02/03/1952	15	3	\$57,000	\$27,000	98	144	0					
2									0					
3									0					
4									0					
5									0					
6									0					
7									0					
8									0					
9									0					
10									0					
11									0					
12									0					
13									0					
14	14 f	02/26/1949	15	1	\$35,100				1					
15	15 m	08/29/1962	12	1	\$27,300				0					
16	16 m	11/17/1964	12	1	\$40,800				0					
17	17 m	07/18/1962	15	1	\$46,000	\$14,250	97	48	0					
18	18 m	03/20/1956	16	3	\$103,750	\$27,510	97	70	0					
19	19 m	08/19/1962	12	1	\$42,300	\$14,250	97	103	0					
20	20 f	01/23/1940	12	1	\$26,250	\$11,550	97	48	0					

일원배치 분산분석

종속변수(E): 현재급여 [현재급여]

요인(F): 직종구분 [직종]

통계량

- ☒ 기술통계(D)
- ☐ 모수 및 변량효과(F)
- ☒ 분산 동질성 검정(H)
 - ☐ Brown-Forsythe
 - ☐ Welch
- ☐ 평균 도표(M)

결측값

- ☒ 분석별 결측값 제외(A)
- ☐ 목록별 결측값 제외(L)

확인 불여넣기(P) 재설정(R) 취소 도움말

IBM SPSS Statistics 프로세서 준비 완료

Unicode:ON



제4절 집단간 차이분석은 어떻게 실행하는가?

[그림 12-14] 일원분산분석 창 2

Employee data.sav [데이터세트1] - IBM SPSS Statistics Data Editor



표시: 10 / 10 변수

	번호	성별	생년월일	피교육	직종	현재급여	최초급여	근무월수	경력	소수민족	변수	변수	변수	변수	변수
1	1 m		02/03/1952	15	3	\$57,000	\$27,000	98	144	0					
2						\$18,750	\$18,750	98	36	0					
3						\$12,000	\$12,000	98	381	0					
4						\$13,200	\$13,200	98	190	0					
5						\$21,000	\$21,000	98	138	0					
6						\$13,500	\$13,500	98	67	0					
7						\$18,750	\$18,750	98	114	0					
8						\$9,750	\$9,750	98	0	0					
9						\$12,750	\$12,750	98	115	0					
10						\$13,500	\$13,500	98	244	0					
11						\$16,500	\$16,500	98	143	0					
12						\$12,000	\$12,000	98	26	1					
13						\$14,250	\$14,250	98	34	1					
14	14 f		02/26/1949	15	1	\$35,100	\$16,800	98	137	1					
15	15 m		08/29/1962	12	1	\$27,300	\$13,500	97	66	0					
16	16 m		11/17/1964	12	1	\$40,800	\$15,000	97	24	0					
17	17 m		07/18/1962	15	1	\$46,000	\$14,250	97	48	0					
18	18 m		03/20/1956	16	3	\$103,750	\$27,510	97	70	0					
19	19 m		08/19/1962	12	1	\$42,300	\$14,250	97	103	0					
20	20 f		01/23/1940	12	1	\$26,250	\$11,550	97	48	0					

일원배치 분산분석

종속변수(E):

현재급여 [현재급여]

요인(F):

직종구분 [직종]

확인 불여넣기(P) 재설정(R) 취소 도움말

사원코드 [번호]
 생년월일 [생년월일]
 피교육년수 [피교육]
 최초급여 [최초급여]
 현 근무월수 [근무월수]
 입사전 타 근무월수 [...]
 소수민족 여부 [소수...]

대비(N)...
 사후분석(H)...
 옵션(O)...
 붓스트랩(B)...

데이터 보기 변수 보기



제4절 집단간 차이분석은 어떻게 실행하는가?



[그림 12-15] 일원분산분석 결과 창

*출력결과2 [문서2] - IBM SPSS Statistics Viewer

파일(F) 편집(E) 보기(V) 데이터(D) 변환(T) 삽입(I) 형식(O) 분석(A) 다이렉트 마케팅(M) 그래프(G) 유틸리티(U) 확장(X) 창(W) 도움말(H)



- 출력결과
 - 로그
 - 일원배치 분산분석
 - 제목
 - 노트
 - 기술통계
 - 분산의 동질성 검정
 - ANOVA
 - 사후검정
 - 제목
 - 다중비교
 - 동질적 부분집
 - 제목
 - 현재급여

일원배치 분산분석

기술통계

현재급여

	N	평균	표준편차	표준오차	평균에 대한 95% 신뢰구간		최소값	최대값
사무직	363	\$27,838.54	\$7,567.995	\$397.217	\$27,057.40	\$28,619.68	\$15,750	\$80,000
관리직	27	\$30,938.89	\$2,114.616	\$406.958	\$30,102.37	\$31,775.40	\$24,300	\$35,250
경영직	84	\$63,977.80	\$18,244.776	\$1,990.668	\$60,018.44	\$67,937.16	\$34,410	\$135,000
전체	474	\$34,419.57	\$17,075.661	\$784.311	\$32,878.40	\$35,960.73	\$15,750	\$135,000

분산의 동질성 검정

현재급여

Levene 통계량	자유도1	자유도2	유의확률
59.733	2	471	.000

ANOVA

현재급여

	제곱합	자유도	평균제곱	F	유의확률
집단-간	8.944E+10	2	4.472E+10	434.481	.000



제5절 분산분석과 다른 통계분석기법간의 관계

〈표 12-10〉 근원척도와 통계분석기법

독립변수의 근원척도	종속변수의 근원척도	통계분석기법
명목척도 등간척도/비율척도* 등간척도/비율척도	등간척도/비율척도 등간척도/비율척도 명목척도	분산분석 회귀분석 판별분석, logistic 회귀분석

* 명목척도 사용가능