



통계기반 데이터 분석 (Ch 7)

서강대학교 정보통신대학원 데이터사이언스 & A.I 정화민 교수



평균비교

모집단이 두 개인 경우

- 두 모집단의 종류
- 서로 독립인 두 집단에서의 평균 차이 검정
- 서로 대응인 두 집단에서의 평균 차이 검정

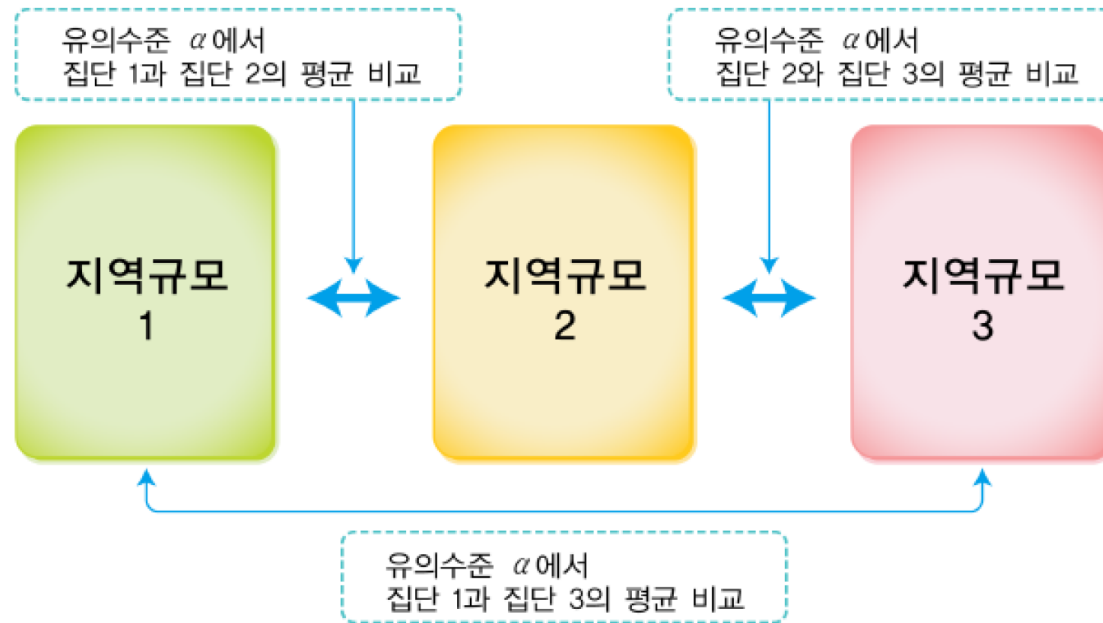
모집단이 세 개 이상

- 분석방법 : 일원분산분석
- 분산분석표

분산분석

모집단이 세 개 이상

- 모집단이 세 개 이상일 경우
 - 서로 독립인 모집단의 개수가 3개 이상으로 확장한 경우
다음 그림과 같이 모집단이 세 개일 때 독립인 두 모집단의 평균 비교를 2개씩 짝을 지어 비교하는 경우



모집단이 세 개 이상

- 유의수준을 0.05로 하여 각각의 평균비교를 통해 차이가 발생하는 집단을 찾을 수 있음.

- 일원배치 분산분석

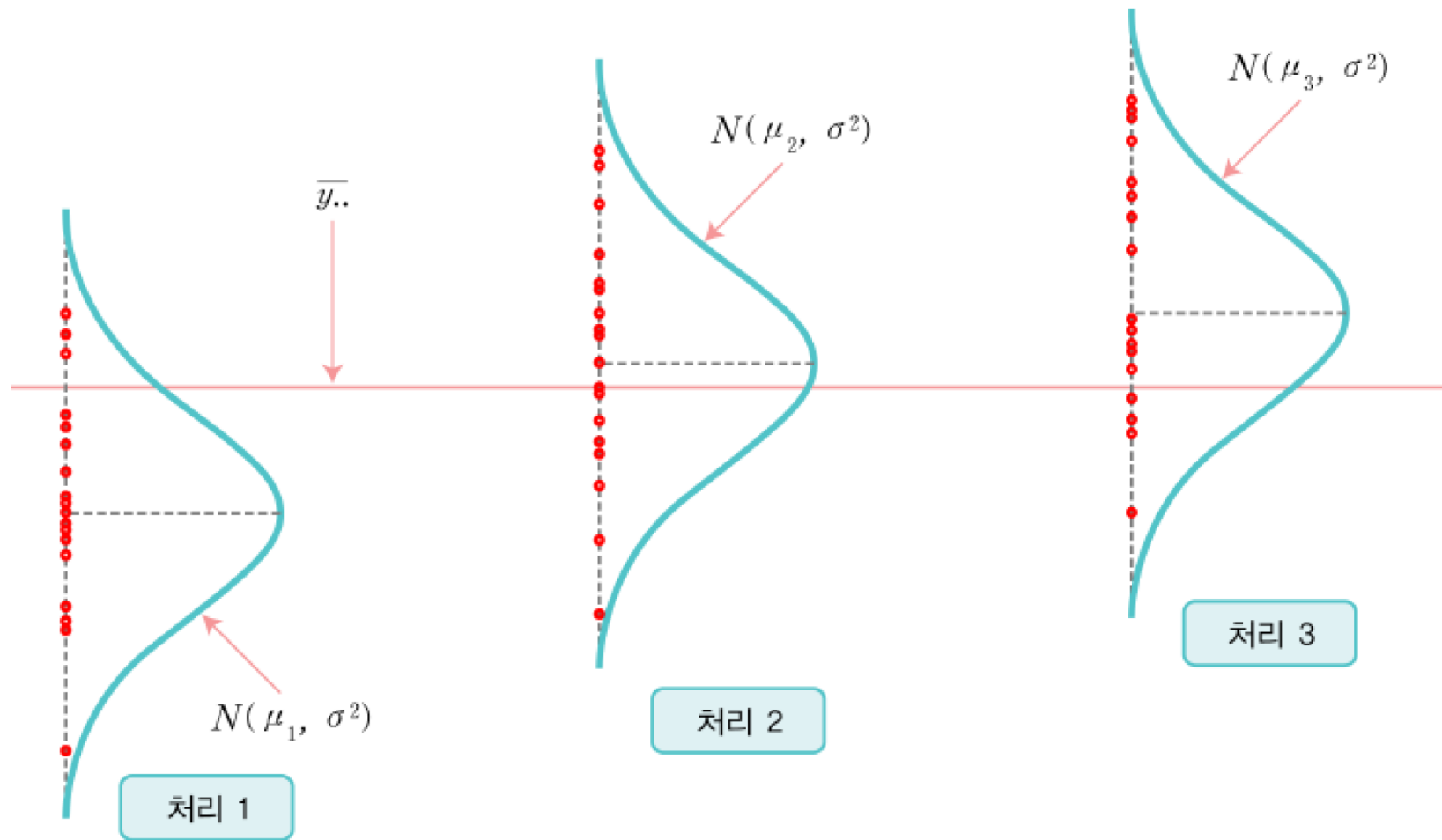
- 집단을 구분하는 요인이 하나에 대해

- 예) 앞선 그림에서 사용된 지역규모 : 집단을 구분하는 요인으로 3개의 수준.
- 자료의 변동(분산)이 발생한 과정을 분석하여 요인에 의한 변동(분산)과 요인을 통해 나누어진 각 집단 내의 변동(분산)을 구함.
- 요인에 의한 변동이 의미 있는 크기를 가지는지를 검정함($p < 0.05$)

- 지역규모별 나이의 차이에 대한 평균 검정을 통해 일원배치분산분석(One way ANOVA)의 과정

분산분석

모집단이 세 개 이상



Source : 제대로 알고 쓰는 R 통계분석

분산분석

모집단이 세 개 이상

■ 가설 수립

- 영가설(귀무가설) : “모든 처리의 평균이 (전체의 모평균과) 같다.” 혹은 “각 처리의 효과는 없다.”

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k, \text{ 혹은 } \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$$

- 예) 지역규모별로 응답자의 연령의 평균은 동일하다.
- 대안가설(연구가설) : 평균의 차이가 있는 것으로 “적어도 한 개의 처리의 평균은 다르다.” 혹은 “적어도 한 개의 처리는 효과가 있다.”
 $H_1 : \text{적어도 하나의 } \mu_i \text{ 는 다르다, 혹은 적어도 하나의 } \alpha_i \neq 0 \text{ 이다.}$
- 예) 지역규모별로 응답자의 연령의 평균은 차이가 있다. (적어도 한 집단의 평균은 차이가 난다)

국내 대학생의 취업태도 및 취업준비 비용에 관한 연구*

정범석* · 정화민**

<요 약>

본 연구는 청년실업의 증가와 이에 대한 대학생들의 취업준비비용, 취업태도에 대한 구체적인 연구가 부족한 상황에서 심각한 취업난에 치해있는 국내 대학생들의 취업을 위한 비용지출의 실태와 취업효능감, 취업태도, 대학교 교육여건 인식에 대한 연구이다.

이를 위하여 국내 4년제 대학교와 전문대생과의 취업비용과 취업태도, 교육여건, 자기효능감의 차이와 인과관계를 통계적으로 알아봄으로써 대학에서의 학생 취업관련 정책과 업무에 시사점을 제공하고자 한다.

교육훈련의 경우 전문대생은 자격증취득, 4년제 대학생은 토익 등 영어교육을 우선시 하는 것으로 조사되었고, 전문대생과 4년제 대학생의 취업비용의 차이분석에서는 해외연수비용이 전문대생은 평균 809만원, 4년제 대학생은 1,559만원, 취업을 위한 교육훈련비용도 4년제 대학생이 200만원, 전문대생은 113만원, 의류장화비용도 전문대생은 66만원, 4년제 대학생은 97만원으로 4년제 대학생이 더 많은 비용을 지출 하는 것으로 나타났다.

전공계열별 해외연수비용과 의류장화비용의 차이검증에서는 컴퓨터경영계열, 전기전자계열, 디자인계열, 사회계열에서 사회계열 학생의 해외연수비용과 의류장화비용이 가장 높은 것으로 조사되었다.

성별에 따른 대학생의 취업비용의 차이도 치아교정비용에서 남학생보다 여학생이 더 많이 지출하는 것으로 나타났다. 대학생 가정의 월평균 수입에 따른 대학생의 취업비용차이는 해외연수비용, 교육훈련비용, 의류장화비용에서 월 소득이 많은 가정의 대학생들이 그렇지 못한 가정의 학생들보다 비용을 더 지불하는 것으로 나타났다. 취업태도는 전문대생이나 4년제 대학생이 인지하는 것은 차이가 없는 것으로 나타났으나, 취업효능감, 학교의 교육여건 인식의 경우 전문대생보다 4년제 대학생이 통계적으로 더 긍정적으로 인식하는 것으로 나타났다.

본 연구의 결과를 통해 취업준비비용에서 해외어학연수비용이 가장 큰 금액을 차지하고 있으며, 대학유형별 지출의 차이도 있는 것으로 나타났다. 대학유형별 취업효능감, 학교교육여건의 인식도 4년제 대학생이 전문대생보다 더 높게 조사되었다. 자기효능감과 교육여건은 취업태도에 긍정적 영향을 미치고 있고, 대학에서의 취업태도를 긍정적으로 하기 위한 프로그램의 적용과 전문대학은 자격증 교육, 4년제 대학은 영어 프로그램에 대한 지원과 프로그램 선정에 정확적 시사점을 제시하였다.

핵심주제어: 대학생 취업준비비용, 취업태도, 취업효능감, 교육여건, 대학생 해외어학연수비용, 대학유형별 대학생 취업준비비용

<표 1> 조사대상의 일반적 특성

		빈도	퍼센트
대학	전문대학	300	62.0
	4년제 대학	184	38.0
전공계열	컴퓨터 경영계열	134	27.7
	전기전자계열	59	12.2
	디자인계열	67	13.8
	사회계열	148	30.6
	기타	76	15.7
학년	1학년	67	13.8
	2학년	166	34.3
	3학년	165	34.1
	4학년	70	14.5
	기타	16	3.3
성별	합계	484	100.0
	남자	260	53.7
	여자	224	46.3
월평균 수입	합계	484	100.0
	200만원 이하	106	21.9
	300만원 전후	187	38.6
	400만원 전후	83	17.1
	500만원 이상	87	18.0
	무응답	21	4.3
	합계	484	100.0
그룹별 연령	21세 이하	146	30.2
	22~24세 이하	197	40.7
	25세 이상	141	29.1
	합계	484	100.0

논문사례

<표 3> 대학유형별 취업비용 차이

(단위 : 만원)

대학		평균	표준편차	t	p
해외연수비용	A	809.92	843.97	-5.16	.000**
	B	1539.87	1241.22		
교육훈련비용	A	113.35	159.60	-3.13	.002**
	B	202.28	297.40		
컨설팅 비용	A	81.07	459.12	.158	.874
	B	73.96	107.32		
치야교정	A	445.67	417.06	1.36	.175
	B	366.28	271.59		
피부	A	119.31	270.19	-.396	.693
	B	134.87	185.22		
성형	A	666.31	1397.77	.698	.487
	B	535.26	787.21		
의류잡화비용	A	66.96	75.11	-3.19	.002**
	B	97.98	92.99		

* : $p < 0.05$, ** : $p < 0.01$

A : 전문대, B : 4년제대학

<표 4> 전공계열별 취업준비비용 차이

(단위 : 만원)

		N	평균	표준편차	F/p	scheffe
해외 연수	A	55	1057.8	922.2	5.022/ .001**	b,c < d,e
	B	24	660.5	780.8		
	C	23	764.5	776.6		
	D	75	1481.4	1296.3		
	E	53	1557.5	1241.6		
	계	230	1240.3	1149.4		
의류 잡화	A	98	68.1	55.1	4.219/ .002**	a,b < d
	B	36	49.1	43.0		
	C	30	91.2	111.4		
	D	85	106.7	109.7		
	E	51	73.1	74.8		
	계	300	79.9	84.3		

* : $p < 0.05$, ** : $p < 0.01$

A : 컴퓨터경영계열 B : 전기전자계열 C : 디자인계열 D : 사회계열 E : 기타

분산분석

모집단이 세 개 이상

```
1  PlantGrowth=read.csv("PlantGrowth.csv")
2  PlantGrowth
3  with(PlantGrowth,tapply(weight,group,mean))
4  with(PlantGrowth,tapply(weight,group,sd))
5  boxplot(weight~group,col="red", data=PlantGrowth)
6  out=lm(weight~group, data=PlantGrowth)
7  out
8  lm(formula=weight~group, data=PlantGrowth)
9  #분산분석표
10 anova(out)
11 #분산의 동질성 검정(등부산가정됨)
12 shapiro.test(resid(out))
13 # dunnett(이분산)
14 install.packages("multcomp")
15 library(multcomp)
16 out=lm(weight~group,data=PlantGrowth)
17 dunnett=glht(out,linfct=mcp(group="Dunnett"))
18 summary(dunnett)
19 plot(dunnett)
20 #Tukey(등분산)
21 tukey=glht(out,linfct=mcp(group="Tukey"))
22 summary(tukey)
23 plot(tukey)
```

Logistic Regression 사례



이대식 한국전자통신연구원 진단치료기연구실 책임연구원

**(인공지능) 기계 학습법을 통해서 데이터를 축적하게 되고요,
이 데이터를 비교함으로써 폐암 환자와 정상인을 구분하게 됩니다.**

Source <https://www.youtube.com/watch?v=Kqx6JVLygcI>

Logistic Regression Analysis

로지스틱 회귀(logistic regression)는 D.R.Cox가 1958년 에 제안한 확률 모델로서 독립 변수의 선형 결합을 이용하여 사건의 발생 가능성을 예측하는데 사용되는 통계 기법이다.

로지스틱 회귀는 종속변수가 이항형 문제(즉, 유효한 범주의 개수가 두개인 경우)를 지칭할 때 사용된다. 이외에, 두 개 이상의 범주를 가지는 문제가 대상인 경우엔 다항 로지스틱 회귀 (multinomial logistic regression) 또는 분화 로지스틱 회귀 (**polytomous logistic regression**)라고 하고 복수의 범주이면서 순서가 존재하면 서수 로지스틱 회귀 (ordinal logistic regression) 라고 한다.^[2] 로지스틱 회귀 분석은 의료, 통신, 데이터마이닝과 같은 다양한 분야에서 분류 및 예측을 위한 모델로서 폭넓게 사용되고 있다.

source : 위키백과

Logistic Regression Analysis

– 로지스틱 회귀분석은 일반적인 선형회귀분석이 수치예측을 위해 활용되는 것과는 달리, 예측하고자 하는 목표변수(y)의 범주를 분류하고자 할 때 사용한다. 즉, 예측하고자 하는 것이 목표변수 Y 값이 아니라, 목표변수 Y 가 특정 범주(i)가 될 확률 $P(Y=i)$ 이다.

1) 로지스틱 함수 (logistic function)

로지스틱 회귀분석을 위해서는 우선 예측하고자 하는 목표변수 Y 의 범주가 0,1 두 가지만 있다고 할 때 (이항 로지스틱 회귀모델을 가정) 목표변수 Y 의 범주가 1이 될 확률 $p(Y=1) = P(Y)$ 로 표기

$$\frac{P(Y)}{1 - P(Y)} = \exp(\beta_0 + \beta_1 X)$$

Logistic Regression Analysis

이분형(종속변수가 1,0) 로지스틱의 분석 사례는 의료분야 죽느냐(1) 사느냐(0) ?
마케팅 분야의 구매를 한다(1), 하지 않는다(0) 등 이분형으로 나뉘어질 확률을
계산함 (종속변수의 코딩이 1, 0으로 되면 이벤트가 있는 쪽을 1로 코딩함)

장점	단점
<ul style="list-style-type: none">· 선형통계모형의 이론에 기반한 정교하고 체계적인 모수 추정이 가능하다.· 확률모형이므로, 목표변수의 범주 확률값을 추정할 수 있다.· 추정된 모형의 계수에 대한 해석이 가능하며, 독립변수들의 유의성 및 영향력 등 결과 분석 시 유용한 해석이 가능하다.	<ul style="list-style-type: none">· 데이터 세트의 차원이 매우 많을 때 모형의 추정 정확도가 다른 머신러닝 기법에 비해 좋지 않다.· 추정 방법상 x값이 매우 커지거나 작아지면 확률값이 1(혹은 0)에 매우 가까워져서 수치계산 정확도가 떨어지게 되며, 반복 계산 시 오버피팅이 빈번하게 발생한다.· 복잡한 비선형적 분류가 필요한 경우에는 분류 정확도가 좋지 않다.

Logistic Regression Analysis 활용분야

로지스틱 회귀분석은 일반적으로 금융권의 신용평가모형이나 이탈예측모형 등에서 전통적으로 많이 활용되어 왔다. 그 외에도 구매예측이나 마케팅 반응 예측 등 수치 데이터를 활용한 분류 예측 문제에 많이 활용되고 있다.

Logistic Regression Analysis

The screenshot shows the DBpia search results page. The top navigation bar includes the DBpia logo, search filters (문헌포함, 저자명, 간행물명, 발행기관명), and buttons for search, 상세검색, 지식공유 캠페인, and 간편외부접속. Below the navigation bar, there are tabs for 주제분류, 간행물, 발행기관, 저자, 개인 맞춤 추천, and 마이페이지. The main content area displays the search results for the keyword '송영숙(송실대학교)'. The results show a list of documents, with the first one being '대학 재학 중 근로 경험이 취업에 미치는 영향 분석 [KCI등재]' by 송영숙. The document is available in PDF, QuickView, and TextView formats. The left sidebar contains filters for 검색결과, 결과내 재검색, 검색어, 검색, 검색결과와 좀더보기, 발행연도, 주제분류, 간행물명, 자료유형, and 해외 등재 정보.

DBpia

문헌포함 송영숙 goms

저자명 간행물명 발행기관명

검색 상세검색

> 지식공유 캠페인

> 간편외부접속 ?

주제분류 간행물 발행기관 저자 개인 맞춤 추천 마이페이지

홈 > 검색결과

Q 검색결과

결과내 재검색

전체 검색

검색결과와 좀더보기

▼ 발행연도

선택 선택 검색

• 1981년 (1)

▼ 주제분류

• 사회과학 (1)

▼ 간행물명 가나다 ↓

• 한민족어문학 (1)

▼ 자료유형 ?

• 전자저널 논문 (1)

▼ 해외 등재 정보

• 해당없음 (1)

검색결과 △ 검색결과 알림 서비스 신청

오늘 본 검색어 검색환경 설정

키워드 관련 저자 ? 저자명을 클릭하시면, '저자페이지'로 이동합니다.

송영숙(송실대학교)

검색어 : 전체='송영숙 goms'

자료건수 : 총 1개

문헌 검색 본문포함 본문제외 정렬방식 유사도순 20개

전체 구매하기 논문모두저장 보관함 인용하기

이용가능 구매가능 이용불가

대학 재학 중 근로 경험이 취업에 미치는 영향 분석 [KCI등재]

송영숙

한국산학기술학회, 한국산학기술학회 논문지 18(2), 2017.2, 287-293 (7 pages)

원문저장 PDFView QuickView TextView 상세보기

이용수 68

피인용수 0

DBpia 추천논문 >

학 교 전 자 도 서 관 으 로
들 어 가 그 림 의 논 문 을 다 운
받 습 니 다 .

(GOMS 공공자료를 이용한
로지스틱 회귀분석방법을
이용하였습니다.)

logistic Regression Analysis

Table 1. Characteristics of Respondents

Spec.		N(%)
Gender	Male	9528(52.5)
	Female	8632(47.5)
Field of Study	Humanities	2087(11.5)
	Society	3524(19.4)
	Education	1513(8.3)
	Engineering	4986(27.5)
	Natural Science	2290(12.6)
	Medical Science	1151(6.3)
	Art and Sports	2609(14.4)
Institution Type -Ownership	National	3573(19.7)
	Public	171(0.9)
	Private	14412(79.4)
	Other	4(0.0)
Institution Type -Year	2 Year	5395(29.7)
	4 Year	12325(67.9)
	University of Education	440(2.4)
Institution Area	Seoul	4075(22.4)
	Kyunggi	4711(25.9)
	Chungcheong	2904(16.0)
	Youngnam	4299(23.7)
	Honam	2171(12.0)
All		18160(100.0)

Table 3. Difference in Employment and Permanent Employment according to Internship

Spec.		Internship		χ^2/p
		Yes	No	
Employment	Yes	426(2.3)	12236(67.4)	4.200*
	No	153(0.8)	5345(29.4)	/.044
Permanent employment	Yes	342(1.9)	9218(50.8)	9.901**
	No	237(1.3)	8386(46.1)	/.002
All		579(3.2)	17581(96.8)	

*p< .05, **p< .01, ***p< .001

인구통계적 특성이
빈도분석으로 잘 나타나
있습니다.

원 데이터는 GOMS 2013년
데이터 입니다.

재학중 근로경험유무, 인턴경험유무의
집단간 **분포의 차이**를 카이제곱
검정으로 봤습니다.

Logistic Regression Analysis

Table 5. The Effect of Internship and The Relatedness of Work Experience and Employment on College Graduates' Permanent Employment

Variables	B	S.E.	Wald	p	Exp(B)
Internship	.003	.015	.048	.826	1.003
The Relatedness of Work Experience and Employment	.199	.088	5.172	.023*	1.220
Constants	.158	.039	16.307	.000	1.171

*p< .05, **p< .01, ***p< .001

관련전공 일경험유무는 유의 확률 값이 .023으로 유의수준 .05 미만으로 나타나 정규직 취업에 통계적으로 유의한 영향이 있는 것으로 나타났다. 그러나 인십턴경험은 정규직 취업에 통계적으로 유의한 영향을 주지 않는 것으로 나타났다.

(관련전공 일경험한 사람이 그렇지 못한 사람에 비하여 정규직으로 갈 확률이 1.22배 높다는 것을 알 수 있다.