

# 벤처 기업의 매출 규모에 영향을 주는 요인 분석

Author : Ji Yoon Kim, Ji Hoo Kim, Myungjin Woo

## ABSTRACT

정부에서는 다양한 정책으로 벤처기업을 육성하기 위해 노력하고 있다. 벤처기업의 성장을 위해서는 벤처기업의 매출에 영향을 미치는 요인을 아는 것이 중요하다. 2020년 2500개의 한국 벤처 기업데이터에서 실무연수, 정규직, 대졸인력, 산업재산권, 기업성장단계, 부채, 지분구조, 시장점유율, 총 8개의 변수를 사용했다.

단회귀 분석, 다중 회귀 분석, 비선형 회귀분석을 차례로 진행하여 결론적으로 유의미한 변수만을 포함한 최종모델은  $\text{규모} = \text{정규직} + \text{대졸인력} + \text{기업성장단계} + \text{부채} + \text{지분구조} + \text{시장점유율} + \text{정규직} * \text{대졸인력}$ 이다. 창업자 실무경험과 같은 개인적인 경험보다 기업의 성장단계, 정규직 수, 부채, 시장 점유율과 같은 기업의 구조와 활동이 매출에 유의미한 영향을 미쳤다.

본 논문은 벤처기업의 성장을 위해 회귀분석을 사용해 벤처기업의 성공요인을 찾고 매출액을 예측할 수 있는 회귀모델을 제시했다는 것에 의의가 있다.

## I. Introduction

### 1. 주제 선정 이유

벤처 기업이란 첨단 기술과 아이디어를 가진 사업자가 높은 기대수익을 확신하고 아무도 시작하지 않은 새로운 사업을 위해 설립하는 기업을 말한다. 기업의 특성 상 다른 종류의 기업들보다 성장성이 큰 반면 누구도 시도하지 않았기 때문에 여러 위험 요소로 인해 실패로 끝날 확률이 더욱 높다. 그렇기에 벤처 기업들이 성공하기 위해서 어떤 요인이 중요한지 알고 이에 따른 준비와 대책이 필요하다. 성공의 기준은 다양하지만 그 중 본 연구에서는 벤처 기업의 성공의 척도를 매출 규모로 잡고 어떠한 요소들이 벤처 기업을 실패가 아닌 성공으로 이끄는 지 분석하는 것이 목표이다.

## 2. 선행 연구 검토

벤처 기업의 성공과 관련된 선행연구는 다양한 방식으로 진행되어 왔다. 이설빈(2017)은 벤처창업 성공요인과 관련된 연구를 진행하여 가장 중요한 것은 창업가정신이라는 결론을 도출하였다. 그 다음으로 혁신성과 경제성, 기술성 요인 순으로 벤처기업의 성공과 유의미한 관련이 있다. 이는 일반 창업에서 중요시되는 창업자의 기업가 정신이 벤처창업에서도 공통적으로 적용된다는 것을 알 수 있다. 그 뿐만 아니라 벤처 기업의 특성에 따라 긍정적인 창업동기와 더불어 창조적인 기술 활용이 이루어질 때 벤처 기업이 성공할 수 있다는 결론을 내렸다. 김인수 등(2017)은 연구 주제에 기업가정신과 혁신활동이라는 두가지 요인이 벤처기업의 성공에 가장 중요하다고 생각하였다. 그리하여 이 두가지 요인에 가장 큰 영향을 미치는 요인들에 대한 연구를 진행하였다. 연구의 결과로 기업가 정신에는 혁신성이 가장 중요하다는 결론을, 그리고 기술상업화는 기술혁신에 의해서 가장 크게 영향을 받는다는 결론을 내리게 되었다. 국내 기업의 5년 생존율이 약 30%된다는 사실에 의하여 벤처기업의 지속가능성에 주목한 연구도 있었다. 안원영(2018)은 지속 가능한 기업으로의 성장을 위해서 성공 요인을 ‘인적 자본’, ‘사회적자본’, ‘재무적자본’으로 나누어 살펴보았다. 인적자본의 독립변수는 제품개발능력, 리더십, 인재 확보능력이고, 사회적자본의 독립변수는 구조적차원과 인지적차원이라는 것을 도출해냈다. 또한 재무적 자본에 자체조달자본규모가 유의미한 영향을 미친다는 결론을 내렸다.

벤처기업과 관련된 선행연구들은 벤처기업이 지속적으로 성공하기 위한 초석을 제공해주고 있다. 특히나 창업가가 지녀야 할 마인드와 지속적인 투자지원이라는 정책도 제시한다는 점에서 연구 의의가 있다. 그러나 대부분의 선행연구가 창업가 정신이 가장 중요하다는 요인이라고 염두해둔 것과 달리 본 연구는 벤처기업의 성공을 ‘매출’이라는 구체적인 변수로 설정하여 진행하였다. 또한 규모, 기업성장단계, 지분구조 등등의 독립변수들로 단순회귀분석부터 비선형회귀분석까지 시도해보며 최적의 모델을 구축해 나간다는 점에서 의의가 있다.

## II. Model

### 1. 데이터 소개

가장 최근 데이터인 2020년의 벤처기업 현황을 분석할 수 있는 총 307개의 변수를 포함한 2500개의 데이터를 활용하였다. 이 중 벤처기업의 매출 규모에 직접적인 영향을 끼칠 것으로 보이는 실무년수, 정규직, 대출인력, 산업재산권, 기업성장단계, 부채, 지분구조, 시장점유율 변수들을 추출해 분석에 이용할 것이다.

구분_법인/가중치	[벤처]	지역[벤처]	업종[벤처]	규모A1-4_벤처A1-5_회사A1-6_KSIC[벤처]	업력A2-2_기업A2-3_제조A2-3_일부B2-1_창업B2-2_창업B2-3_창업B2-3-1_창B2-4_창업
개인	0.079104	1	8	3	4 경기 62021 3 3 4 1 1 3 3 10
개인	0.079104	3	8	3	2 경남 62021 3 4 1 1 2 4 1 6
개인	0.431785	1	10	3	1 경기 1299 2 3 1 1 2 5 6
개인	0.431785	3	10	3	2 경남 42499 3 4 4 1 2 3 1 6
개인	0.450524	2	10	2	2 강원 42412 3 4 1 1 2 3 1 6
개인	0.450524	4	10	2	2 대구 42201 4 4 4 1 2 3 1 5
개인	0.512647	1	5	6	2 경기 29192 4 3 2 30 1 3 3 1 6
개인	0.512647	1	5	6	2 인천 29299 4 4 2 10 1 3 3 1 6
개인	0.512647	3	5	6	2 부산 29120 3 4 2 10 1 3 4 1 6
개인	0.512647	3	5	6	2 부산 29132 4 4 2 25 1 2 5 6
개인	0.512647	4	5	6	2 경북 30332 4 5 1 1 1 3 1 6
개인	0.539731	1	10	3	1 서울 58113 4 4 2 25 1 2 3 4 8
개인	0.539731	1	10	3	2 서울 58122 4 4 1 2 1 2 1 1
개인	0.539731	2	10	3	2 대전 42311 4 4 4 1 2 1 1 6
개인	0.539731	4	10	3	2 경북 1231 2 4 4 1 2 3 4 6
개인	0.63518	2	10	4	1 대전 1299 2 3 1 1 2 3 3 9
개인	0.672358	1	9	4	2 경기 18119 4 4 1 1 2 5 6
개인	0.672358	4	9	4	1 대구 18119 3 4 1 1 1 3 1 6
개인	0.711906	1	9	5	1 경기 46315 2 3 2 20 1 2 4 5 6
개인	0.711906	1	9	5	1 경기 46413 2 2 3 2 2 5 6
개인	0.711906	1	9	5	2 인천 18119 4 5 2 30 1 3 5

<sup>1</sup>벤처기업 정밀실태조사(2020)

## 2. 변수 설명

규모: 매출 규모(억)

실무: 창업 당시 창업자의 실무경험(년)

정규직: 정규직 직원의 수(명)

대졸인력: 최종학력이 대학교 졸업 이상인 직원의 수(명)

산업재산권: 기업이 등록보유한 산업재산권 수(건)

기업성장단계: 1-창업기, 2-초기성장기, 3-고도성장기, 4-성숙기, 5-쇠퇴기

부채: 2019년 부채(백만원)

지분 구조: 창업자, 임직원, 가족, 벤처캐피털 및 기관투자자, 엔젤 및 엑셀러레이터, 개인투자자의 비율(%)

시장점유율: 주력 제품 국내, 해외 시장점유율(%)

<sup>1</sup> 출처: MDIS(마이크로데이터 통합서비스)

<https://mdis.kostat.go.kr/ofrData/selectOfrDataDetail.do?survId=14203&itmDiv=2&nPage=3&itemId=2010&itemNm=%EA%B2%BD%EC%A0%9C%EC%9D%BC%EB%B0%98%C2%B7%EA%B2%BD%EA%B8%B0/%EA%B8%B0%EC%97%85%EA%B2%BD%EC%98%81#click>

### 3. 데이터 전 처리

지분 구조와 시장점유율 변수에서 발견된 결측치를 각 변수의 중앙값으로 대체했다. 전처리를 거친 최종 데이터는 아래 표와 같다.

	규모	기업성장단계	실무년수	정규직	대졸인력	지분구조	부채	산업재산권	시장점유율
count	2500.000000	2500.000000	2500.000000	2500.000000	2500.000000	2500.000000	2500.000000	2500.000000	2500.000000
mean	114.850400	3.261200	10.531960	47.598800	21.310800	56.468720	8755.375200	11.918800	4.770828
std	133.526296	0.824892	6.982106	64.395016	32.537438	29.362204	18476.688633	23.439709	9.256763
min	1.000000	1.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	20.000000	3.000000	5.000000	10.000000	4.000000	40.000000	933.000000	1.000000	0.100000
50%	80.000000	3.000000	10.000000	23.000000	10.000000	55.000000	3008.500000	4.000000	1.000000
75%	120.000000	4.000000	15.000000	57.000000	24.000000	80.000000	7883.000000	12.000000	5.000000
max	500.000000	5.000000	40.000000	658.000000	490.000000	100.000000	380473.000000	230.000000	80.000000

### 4. 모델 분석

#### 4.1. Simple Linear Regression

우선 가장 기본적인 단순회귀분석부터 진행하였다. 단순회귀분석을 위해서 고른 독립변수는 ‘정규직’이다. ‘정규직’ 변수는 2019년 정규직의 합계인 변수이다. 이 변수를 선택한 이유는 종속변수인 ‘규모’와 상관계수를 확인해보았을 때 0.67로 가장 높은 상관계수를 가지고 있었기 때문이다.

단순회귀분석을 하기 전에 회귀계수의 분산을 모르기 때문에 Heteroskedastic Robust Standard Error를 적용하여 이분산성을 가정하여 분석을 진행하였다.

단순 회귀 모델을 돌려보았을 때 정규직 변수의 계수는 1.4042으로 나왔다. 그러나 계수가 정말 유의미한 지 결과를 도출하기 위해서 가설 ‘정규직은 매출 규모에 영향을 미치지 않을 것이다.’에 대한 가설 검정을 진행하였다. Z-score는 20.937로 매우 높고 p-value는 0.000로 나와서 유의수준인 0.05 보다 작아서 기각을 할 수 있다. 즉, ‘정규직’ 변수는 매출에 영향을 미치지 않는다는 가설을 기각하고 계수 1.4042가 정말 유의미하다는 것을 도출할 수 있었다.

이 회귀식을 해석하면 정규직이 한 단위 증가하면, 매출 규모는 1.4042 단위 만큼 증가한다는 정규직과 매출 규모 사이의 양의 관계가 있다는 것을 도출할 수 있었다. 그러나 모델의 설명력은 R<sup>2</sup>에 따르면 45.9%이다.

```

                        OLS Regression Results
=====
Dep. Variable:          규모      R-squared:          0.459
Model:                  OLS      Adj. R-squared:       0.458
Method:                 Least Squares      F-statistic:       438.4
Date:                  Tue, 07 Jun 2022      Prob (F-statistic): 8.18e-90
Time:                  04:10:05      Log-Likelihood:    -15016.
No. Observations:      2500      AIC:               3.004e+04
Df Residuals:          2498      BIC:               3.005e+04
Df Model:               1
Covariance Type:       HC0
=====
                        coef      std err          z      P>|z|      [0.025      0.975]
-----
const                48.0107      2.878      16.683      0.000      42.370      53.651
정규직                1.4042      0.067      20.937      0.000      1.273      1.536
=====
Omnibus:              919.297      Durbin-Watson:      0.808
Prob(Omnibus):        0.000      Jarque-Bera (JB):    5344.531
Skew:                 1.629      Prob(JB):            0.00
Kurtosis:             9.379      Cond. No.            99.6
=====

```

Notes:

[1] Standard Errors are heteroscedasticity robust (HC0)

## 4.2. Multiple Linear Regression

단순회귀분석으로는 모델의 설명력이 낮기 때문에 여러 개의 변수를 포함한 다중회귀 분석을 진행해보았다. 다중회귀분석은 모든 변수를 포함한 모델에서 시작하여 기준에 통계치를 개선시키는 변수를 차례로 제거해 나가는 방법인 Backward Elimination(후진제거법)을 선택했다. 그리고 모든 다중회귀분석을 진행할 때 단순회귀분석과 마찬가지로 Heteroskedastic Robust Standard Error를 적용하여 이분산성을 가정하였다.

실무년수, 정규직, 대졸인력, 산업재산권, 기업성장단계, 부채, 지분구조, 시장점유율을 모두 포함한 모델 역시 이분산성을 가정하여 분석을 진행하였다.

후진제거법을 위하여 가장 처음으로는 실무년수, 정규직, 대졸인력, 산업재산권, 기업성장단계, 부채, 지분구조, 시장점유율 8개의 변수를 모두 포함한 다중회귀분석을 진행하였다. 그 결과는 <표1>에서 확인할 수 있다. 회귀 계수는 각각 -0.42, 0.90, 0.07, 0.31, 8.5, 0.002, -0.23, 0.85로 나왔다. 그러나 마찬가지로 회귀 계수가 유의미한 지 알아보기 위해서 가설 검정을 진행하였다. 유의수준 0.05 하에서 대졸인력을 제외한 모든 변수들이 유의미하다는 결론이 나왔다. 대졸인력 변수는 p-value가 0.7로써 유의수준 0.05 하에서 기각이 불가능하여 유의미하지 않은 변수이다. 모델의 설명력을 보기 위해서 R2를 확인하여야 하는데, R2는 계수가 많아질수록 자연스럽게 값이 높아진다는 특성이 있다. 그렇기 때문에 다중회귀모델에서는 변수의 계수만큼 패널티를 주는 Adjusted R2를 통해 모델의 설명력을 확인하여야 한다. 이 모델의 Adjusted R2는 0.525로 이전의 단순 회귀 분석을 진행했을 때보다 독립변수가 종속변수의 분산을 잘 설명한다는 것을 알 수 있었다.

‘대졸인력’ 변수는 유의미하지 않기 때문에 Backward Elimination을 적용하여 ‘대졸인

력’을 제외한 다중회귀분석을 진행하였다.

‘대출인력’을 제외한 다중회귀분석 결과는 <표2>와 같다. 실무년수, 정규직, 산업재산권, 기업성장단계, 부채, 지분구조, 시장점유율의 회귀 계수는 -0.4, 0.9, 0.3, 8.4, 0.002, -0.23, -0.86이다. Hypotheses Test를 진행하여 변수가 유의미한지 확인하였다. Significance level 0.05 수준 하에서 실무년수 변수의 p-value 값이 0.06으로 유의미하지 않다는 결론이 나왔다. 이 모델은 Adjusted R2 값이 0.525로 앞의 모델과 동일한 설명력을 갖는다는 것을 알 수 있다.

OLS Regression Results

Dep. Variable:	규모	R-squared:	0.526
Model:	OLS	Adj. R-squared:	0.525
Method:	Least Squares	F-statistic:	152.4
Date:	Mon, 06 Jun 2022	Prob (F-statistic):	4.25e-165
Time:	13:05:07	Log-Likelihood:	-14849.
No. Observations:	2500	AIC:	2.971e+04
Df Residuals:	2493	BIC:	2.975e+04
Df Model:	6		
Covariance Type:	HCO		

	coef	std err	z	P> z	[0.025	0.975]
Intercept	29.9727	8.480	3.535	0.000	13.353	46.592
정규직	0.9399	0.095	9.905	0.000	0.754	1.126
산업재산권	0.3085	0.136	2.262	0.024	0.041	0.576
기업성장단계	8.6713	2.205	3.933	0.000	4.350	12.993
부채	0.0020	0.000	3.988	0.000	0.001	0.003
지분구조	-0.2305	0.075	-3.057	0.002	-0.378	-0.083
시장점유율	0.8055	0.239	3.365	0.001	0.336	1.275

Omnibus:	860.983	Durbin-Watson:	0.902
Prob(Omnibus):	0.000	Jarque-Bera (JB):	5809.615
Skew:	1.460	Prob(JB):	0.00
Kurtosis:	9.874	Cond. No.	1.03e+05

Warnings:

[1] Standard Errors are heteroscedasticity robust (HCO)

[2] The condition number is large, 1.03e+05. This might indicate that there are strong multicollinearity or other numerical problems.

마지막으로 ‘대출인력’, ‘실무년수’를 제외하여 다중회귀분석을 진행하였다.(<표3>) 정규직, 산업재산권, 기업성장단계, 부채, 지분구조, 시장점유율의 회귀 계수는 0.9, 0.3, 8.6, 0.002, -0.23, 0.8이다. Hypotheses test를 진행하였을 때 모든 변수 중 가장 큰 p-value는 0.02로 모든 변수의 회귀 계수가 유의미하다는 결론이 나오게 되었다. 이 모델 또한 Adjusted R2 값이 0.525로 세 모델 모두 동일한 설명력을 갖는다는 결론을 내릴 수 있다. 그리고 모든 변수들이 유의미하기 때문에 더 이상 제거할 변수가 없기 때문에 다중회귀분석을 중단할 수 있다.

세 모델이 모두 동일한 설명력을 가짐에도 불구하고 마지막 모델을 선택해야 한다. 그 이유는 Adjusted R2가 높다고 하여 모든 변수들이 유의미하다는 것을 보장하지 않기 때문이다. 즉, 단순히 모델의 설명력보다 더 중요한 것은 모델에 사용된 독립변수의 논리성과 이론적 근거이기 때문에 모든 변수들이 유의미한 마지막 모델을 선택해야 하는 것이

다.

### 4.3. Nonlinear Regression : Interaction

나아가 선형 모델에서는 나타나지 않는 변수들 간의 연관성을 확인하기 위해 비선형 모델을 돌려보았다. 다양한 비선형 모델 중 변수 간의 Correlation을 알아보기 위해 두 변수 간의 Interaction 을 추가 변수로 선택했다. 실무년수, 정규직, 대졸인력, 산업재산권, 기업성장단계, 부채, 지분구조, 시장점유율 중 상호연관성이 있을 것으로 예측되는 대졸인력\*정규직, 기업성장단계\*시장점유율을 Interaction 변수로 선정했다. 위의 모델들과 마찬가지로 Heteroskedastic Robust Standard Error를 적용하여 이분산성을 가정했음을 밝힌다.

첫 번째로, 실무년수, 정규직, 대졸인력, 산업재산권, 기업성장단계, 부채, 지분구조, 시장점유율, 대졸인력\*정규직, 기업성장단계\*시장점유율을 독립변수로 가지는 비선형 회귀 분석을 진행했다. 결과는 <표4>와 같다. Adjusted R2는 0.543으로 이전의 단순 회귀 분석과 다중 회귀 분석을 진행했을 때보다 독립변수가 종속변수의 분산을 잘 설명한다는 것을 알 수 있었다. 회귀계수는 변수 순서대로 -0.19, 1.07, 0.84, 0.22, 5.77, 0.002, -0.14, 0.39, -0.004, 0.14으로, 우리가 관심있는 정규직\*대졸인력의 회귀계수는 -0.004, 기업성장단계\*시장점유율의 회귀계수는 0.14임을 확인했다.

해당 회귀계수가 유의미한 지 알아보기 위해 가설 검정을 진행했다. 검정 결과, 정규직과 대졸인력의 Interaction의 p-value는 0.003인 반면 기업성장단계와 시장점유율 Interaction의 p-value가 0.472으로 유의수준 0.05 하에서 기각이 불가능하다. 즉, 기업성장단계와 시장점유율은 Interaction이 유의하지 않은 반면, 대졸인력이 많아지면 정규직 인원 또한 따라서 증가한다는 것을 알 수 있다.

따라서 Backward Elimination을 적용하여 전체 변수에서 기업성장단계\*시장점유율만을 제외한 새로운 모델로 비선형회귀분석을 진행했다. 그 결과는 <표5> 같다. 이 모델의 Adjusted R2 값 또한 0.543으로 앞의 모델과 동일한 설명력을 갖는다는 것을 알 수 있다. 회귀 계수는 순서대로 -0.19, 1.07, 0.84, 0.22, 6.42, 0.002, -0.14, 0.85, -0.004이다. 가설 검정을 진행해 유의수준 0.05 수준 하에서 실무년수와 산업재산권 변수의 p-value 값이 각각 0.398과 0.078으로 유의미하지 않다는 결과가 나왔다.

같은 방법으로 실무년수와 산업재산권 변수를 차례로 제거해 최종 모델을 도출할 수 있었다.(실무년수 제거<표6>, 실무년수, 산업재산권 제거<표7>) Adjusted R2는 거의 동일한 0.542으로 설명력을 잃지 않고, 회귀 계수는 정규직,대졸인력, 기업성장단계, 부채, 지분구조, 시장점유율,정규직\*대졸인력 순서로 1.11, 0.85, 7.00, 0.002, -0.15, 0.89, -0.004 임을 확인했다. 가설 검정을 진행했을 때 모든 변수의 p-value 값이 유의수준 0.05를 넘지 않기 때문에 모든 변수의 회귀 계수가 유의미하다고 결론을 내리고 비선형 회귀분석을 중단할 수 있다.



```

=====
                        OLS Regression Results
=====
Dep. Variable:                규모      R-squared:                0.544
Model:                        OLS      Adj. R-squared:           0.542
Method:                       Least Squares      F-statistic:             199.2
Date:                          Mon, 06 Jun 2022      Prob (F-statistic):       4.66e-235
Time:                          15:08:53      Log-Likelihood:           -14802.
No. Observations:              2500      AIC:                     2.962e+04
Df Residuals:                  2492      BIC:                     2.967e+04
Df Model:                      7
Covariance Type:               HCO
=====

```

	coef	std err	z	P> z	[0.025	0.975]
Intercept	19.1178	8.460	2.260	0.024	2.537	35.698
정규직	1.1087	0.132	8.402	0.000	0.850	1.367
대졸인력	0.8530	0.253	3.377	0.001	0.358	1.348
기업성장단계	7.0093	2.162	3.242	0.001	2.772	11.247
부채	0.0019	0.000	4.073	0.000	0.001	0.003
지분구조	-0.1520	0.073	-2.080	0.038	-0.295	-0.009
시장점유율	0.8920	0.238	3.743	0.000	0.425	1.359
정규직:대졸인력	-0.0037	0.001	-3.034	0.002	-0.006	-0.001

```

=====
Omnibus:                      874.151      Durbin-Watson:           0.848
Prob(Omnibus):                 0.000      Jarque-Bera (JB):        4882.049
Skew:                          1.551      Prob(JB):                0.00
Kurtosis:                      9.103      Cond. No.:               1.07e+05
=====

```

#### Warnings:

[1] Standard Errors are heteroscedasticity robust (HCO)  
 [2] The condition number is large, 1.07e+05. This might indicate that there are strong multicollinearity or other numerical problems.

정규직과 대졸인력간의 상관관계가 있어 비선형모델이 구성되었다. Interaction term의 해석은 다음과 같다.

$$\Delta \text{규모} / \Delta \text{정규직} = 1.1087 - 0.0037 * (\text{대졸인력})$$

대졸인력	$\Delta \text{규모} / \Delta \text{정규직}$
0	1.1087
20%	$1.1087 - 0.0037 * 20 = 1.0347$

결론적으로, 최종 비선형 모델을  $\text{규모} = \text{정규직} + \text{대졸인력} + \text{기업성장단계} + \text{부채} + \text{지분구조} + \text{시장점유율} + \text{정규직} * \text{대졸인력}$  로 결론지을 수 있다.

### III. Conclusion

본 연구는 2020년 한국 벤처기업을 대상으로 기업의 매출에 영향을 미치는 요인에 대해서 살펴본 연구이다. 이를 위해서 회귀분석을 실시했다. 종속변수로 매출액을 나타내는 규모를 설정하였고 독립변수로 정규직, 실무년수, 대졸인력, 산업재산권, 기업성장단



계, 부채, 지분구조 그리고 시장점유율을 사용했다.

Regressor	(1)	(2)	(3)	(4)
정규직	1.4042 (0.000) [1.273, 1.536]	0.9026 (0.000)	0.9366 (0.000) [0.751, 1.122]	0.9399 (0.000) [0.754, 1.126]
실무년수		-0.4284 (0.066) [0.885, 0.028]	-0.4426 (0.060) [-0.904, 0.019]	
대졸인력		0.0736 (0.713) [0.318, 0.466]		
산업재산권		0.3100 (0.025) [0.040, 0.580]	0.3045 (0.026) [0.037, 0.572]	0.3085 (0.024) [0.041, 0.576]
기업성장단계		8.4997 (0.000) [4.123, 12.876]	8.4239 (0.000) [4.121, 12.726]	8.6713 (0.000) [4.350, 12.993]
부채		0.0020 (0.000) [0.001, 0.003]	0.0020 (0.000) [0.001, 0.003]	0.0020 (0.000) [0.001, 0.003]
지분구조		-0.2317 (0.002) [-0.379, 0.084]	-0.2334 (0.002) [-0.381, 0.086]	-0.2305 (0.002) [-0.378, 0.083]
시장점유율		0.8455 (0.001) [0.350, 1.341]	0.8264 (0.001) [0.352, 1.301]	0.8055 (0.001) [0.336, 1.275]
Intercept	48.0107	34.9885	35.6815	29.9727
Summary Statistics				
BIC	3.005e+04	2.976e+04	2.976e+04	2.975e+04
Adj. R-squared	0.458	0.525	0.525	0.525
n	2500	2500	2500	2500

다음은 선형모델 4개를 정리한 표이다. 박스 안의 값은 차례대로 Coefficient, (p-value), 신뢰구간[0.025, 0.975]을 의미한다. 선형회귀모델 분석 결과 BIC 2.975e+04, Adjusted R square가 0.525로 모델 (4)가 가장 좋은 설명력을 가졌다.

Regressor	(5)	(6)	(7)	(8)
정규직	1.0736 (0.000) [0.811, 1.336]	1.0750 (0.000) [0.813, 1.337]	1.0748 (0.000) [0.813, 1.337]	1.1087 (0.000) [0.850, 1.367]

실무년수	-0.1946 (0.400) [-0.648, 0.259]	-0.1956 (0.398) [-0.649, 0.258]		
대졸인력	0.8491 (0.001) [0.361, 1.337]	0.8456 (0.001) [0.359, 1.332]	0.8566 (0.001) [0.371, 1.342]	0.8530 (0.001) [0.358, 1.348]
산업재산권	0.2257 (0.082) [-0.028, 0.480]	0.2282 (0.078) [-0.025, 0.482]	0.2298 (0.075) [-0.023, 0.483]	
기업성장단계	5.7796 (0.009) [1.425, 10.135]	6.4284 (0.003) [2.216, 10.641]	6.5283 (0.002) [2.322, 10.735]	7.0093 (0.001) [2.772, 11.247]
부채	0.0018 (0.001) [0.001, 0.003]	0.0018 (0.000) [0.001, 0.003]	0.0018 (0.000) [0.001, 0.003]	0.0019 (0.000) [0.001, 0.003]
지분구조	-0.1418 (0.054) [-0.286, 0.002]	-0.1410 (0.056) [-0.285, 0.003]	-0.1390 (0.059) [-0.283, 0.005]	-0.1520 (0.038) [-0.295, -0.009]
시장점유율	0.3908 (0.517) [-0.792, 1.573]	0.8564 (0.000) [0.389, 1.324]	0.8489 (0.000) [0.384, 1.314]	0.8920 (0.000) [0.425, 1.359]
정규직: 대졸인력	-0.0036 (0.003) [-0.006, -0.001]	-0.0035 (0.003) [-0.006, -0.001]	-0.0036 (0.003) [-0.006, -0.001]	-0.0037 (0.002) [-0.006, -0.001]
기업성장단계: 시장점유율	0.1415 (0.472) [0.244, 0.527]			
Intercept	23.8481	21.6419	19.0100	19.1178
<b>Summary Statistics</b>				
BIC	2.968e+04	2.967e+04	2.967e+04	2.967e+04
Adj. R-squared	0.543	0.543	0.543	0.542
n	2500	2500	2500	2500

다음의 표는 비선형모델 4개를 정리한 표이다. 마찬가지로 박스 안의 값은 차례대로 Coefficient, (p-value), 신뢰구간[0.025, 0.975]을 의미한다. 변수 간의 상관관계를 고려해 Interaction 항을 도입했다. 그 결과 BIC 2.967e+04, Adjusted R square가 0.542로 모델 (8)이 선형, 비선형 모델 모두 통틀어서 가장 좋은 설명력을 가졌다.

따라서 벤처기업의 매출규모를 나타내는 최종모델은 다음과 같다.

**규모 = 정규직+대졸인력+기업성장단계+부채+지분구조+시장점유율+정규직\*대졸인력**

정규직과 대졸인력에는 상관관계가 있어 비선형모델을 구성했다. 벤처기업의 매출 규모에 정규직 및 기업성장단계, 부채와 시장 점유율은 positive 한 영향을 주고 지분구조는 negative한 영향을 주었다. 특히 기업성장단계 변수의 회귀계수가 7.0093로 매우 크기 때문에 벤처기업의 성공에는 기업성장단계가 가장 중요하게 작용한다.

창업자 개인의 실무경험과 같은 개인적인 경험보다 기업의 성장단계, 정규직 수, 부채, 시장 점유율과 같은 기업의 구조와 활동이 매출에 영향을 미쳤다. 따라서 벤처기업의 성공은 개인의 능력이 아니라 기업의 활동과 그에 따른 성과가 중요하다.

본 논문은 벤처기업의 성장을 위해 회귀분석을 사용해 벤처기업의 성공요인을 찾고 매출액을 예측할 수 있는 회귀모델을 제시했다는 것에 의의가 있다.

OLS Regression Results						
=====						
Dep. Variable:	규모	R-squared:	0.527			
Model:	OLS	Adj. R-squared:	0.525			
Method:	Least Squares	F-statistic:	115.1			
Date:	Mon, 06 Jun 2022	Prob (F-statistic):	4.92e-164			
Time:	13:04:39	Log-Likelihood:	-14847.			
No. Observations:	2500	AIC:	2.971e+04			
Df Residuals:	2491	BIC:	2.976e+04			
Df Model:	8					
Covariance Type:	HC0					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
Intercept	34.9885	9.298	3.763	0.000	16.765	53.212
실무년수	-0.4284	0.233	-1.838	0.066	-0.885	0.028
정규직	0.9026	0.138	6.556	0.000	0.633	1.172
대졸인력	0.0736	0.200	0.368	0.713	-0.318	0.466
산업재산권	0.3100	0.138	2.247	0.025	0.040	0.580
기업성장단계	8.4997	2.233	3.806	0.000	4.123	12.876
부채	0.0020	0.000	4.029	0.000	0.001	0.003
지분구조	-0.2317	0.075	-3.081	0.002	-0.379	-0.084
시장점유율	0.8455	0.253	3.346	0.001	0.350	1.341
=====						
Omnibus:	853.675	Durbin-Watson:	0.903			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	5683.208			
Skew:	1.450	Prob(JB):	0.00			
Kurtosis:	9.793	Cond. No.	1.10e+05			
=====						
Warnings:						
[1] Standard Errors are heteroscedasticity robust (HC0)						
[2] The condition number is large, 1.1e+05. This might indicate that there are strong multicollinearity or other numerical problems.						

## 표 2

=====						
Dep. Variable:	규모	R-squared:	0.527			
Model:	OLS	Adj. R-squared:	0.525			
Method:	Least Squares	F-statistic:	131.9			
Date:	Mon, 06 Jun 2022	Prob (F-statistic):	1.62e-165			
Time:	12:18:58	Log-Likelihood:	-14847.			
No. Observations:	2500	AIC:	2.971e+04			
Df Residuals:	2492	BIC:	2.976e+04			
Df Model:	7					
Covariance Type:	HC0					
=====						
	coef	std err	z	P> z	[0.025	0.975]
-----						
Intercept	35.6815	9.056	3.940	0.000	17.933	53.430
실무년수	-0.4426	0.236	-1.880	0.060	-0.904	0.019
정규직	0.9366	0.095	9.893	0.000	0.751	1.122
산업재산권	0.3045	0.137	2.229	0.026	0.037	0.572
기업성장단계	8.4239	2.195	3.837	0.000	4.121	12.726
부채	0.0020	0.000	4.000	0.000	0.001	0.003
지분구조	-0.2334	0.075	-3.097	0.002	-0.381	-0.086
시장점유율	0.8264	0.242	3.416	0.001	0.352	1.301
=====						
Omnibus:	856.998	Durbin-Watson:	0.902			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	5751.491			
Skew:	1.454	Prob(JB):	0.00			
Kurtosis:	9.838	Cond. No.	1.10e+05			
=====						
Warnings:						
[1] Standard Errors are heteroscedasticity robust (HC0)						
[2] The condition number is large, 1.1e+05. This might indicate that there are strong multicollinearity or other numerical problems.						

## 표 3

```

=====
OLS Regression Results
=====
Dep. Variable:      규모      R-squared:      0.526
Model:              OLS      Adj. R-squared:  0.525
Method:              Least Squares      F-statistic:    152.4
Date:                Mon, 06 Jun 2022      Prob (F-statistic):  4.25e-165
Time:                13:05:07      Log-Likelihood:    -14849.
No. Observations:    2500      AIC:              2.971e+04
Df Residuals:        2493      BIC:              2.975e+04
Df Model:             6
Covariance Type:     HC0
=====
               coef      std err          z      P>|z|      [0.025      0.975]
-----
Intercept      29.9727      8.480      3.535      0.000      13.353      46.592
정규직          0.9399      0.095      9.905      0.000      0.754      1.126
산업재산권      0.3085      0.136      2.262      0.024      0.041      0.576
기업성장단계      8.6713      2.205      3.933      0.000      4.350      12.993
부채            0.0020      0.000      3.988      0.000      0.001      0.003
지분구조       -0.2305      0.075     -3.057      0.002     -0.378     -0.083
시장점유율      0.8055      0.239      3.365      0.001      0.336      1.275
=====
Omnibus:          860.983      Durbin-Watson:      0.902
Prob(Omnibus):    0.000      Jarque-Bera (JB):    5809.615
Skew:             1.460      Prob(JB):            0.00
Kurtosis:         9.874      Cond. No.            1.03e+05
=====

Warnings:
[1] Standard Errors are heteroscedasticity robust (HC0)
[2] The condition number is large, 1.03e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

```

## 표4

OLS Regression Results

Dep. Variable:	규모	R-squared:	0.545
Model:	OLS	Adj. R-squared:	0.543
Method:	Least Squares	F-statistic:	143.7
Date:	Mon, 06 Jun 2022	Prob (F-statistic):	7.62e-238
Time:	15:07:41	Log-Likelihood:	-14798.
No. Observations:	2500	AIC:	2.962e+04
Df Residuals:	2489	BIC:	2.968e+04
Df Model:	10		
Covariance Type:	HCO		

	coef	std err	z	P> z	[0.025	0.975]
Intercept	23.8481	9.207	2.590	0.010	5.803	41.893
실무년수	-0.1946	0.231	-0.841	0.400	-0.648	0.259
정규직	1.0736	0.134	8.011	0.000	0.811	1.336
대출인력	0.8491	0.249	3.413	0.001	0.361	1.337
산업재산권	0.2257	0.130	1.741	0.082	-0.028	0.480
기업성장단계	5.7796	2.222	2.601	0.009	1.425	10.135
부채	0.0018	0.000	3.974	0.000	0.001	0.003
지분구조	-0.1418	0.074	-1.926	0.054	-0.286	0.002
시장점유율	0.3908	0.603	0.648	0.517	-0.792	1.573
정규직:대출인력	-0.0036	0.001	-2.977	0.003	-0.006	-0.001
기업성장단계:시장점유율	0.1415	0.197	0.719	0.472	-0.244	0.527

Omnibus:	876.738	Durbin-Watson:	0.852
Prob(Omnibus):	0.000	Jarque-Bera (JB):	4973.053
Skew:	1.551	Prob(JB):	0.00
Kurtosis:	9.174	Cond. No.	1.22e+05

Warnings:

[1] Standard Errors are heteroscedasticity robust (HCO)

[2] The condition number is large, 1.22e+05. This might indicate that there are strong multicollinearity or other numerical problems.

## 표5

OLS Regression Results

Dep. Variable:	규모	R-squared:	0.545
Model:	OLS	Adj. R-squared:	0.543
Method:	Least Squares	F-statistic:	159.0
Date:	Mon, 06 Jun 2022	Prob (F-statistic):	5.85e-238
Time:	15:07:44	Log-Likelihood:	-14798.
No. Observations:	2500	AIC:	2.962e+04
Df Residuals:	2490	BIC:	2.967e+04
Df Model:	9		
Covariance Type:	HC0		

	coef	std err	z	P> z	[0.025	0.975]
Intercept	21.6419	9.063	2.388	0.017	3.879	39.405
실무년수	-0.1956	0.231	-0.845	0.398	-0.649	0.258
정규직	1.0750	0.134	8.046	0.000	0.813	1.337
대출인력	0.8456	0.248	3.404	0.001	0.359	1.332
산업재산권	0.2282	0.129	1.764	0.078	-0.025	0.482
기업성장단계	6.4284	2.149	2.991	0.003	2.216	10.641
부채	0.0018	0.000	3.982	0.000	0.001	0.003
지분구조	-0.1410	0.074	-1.914	0.056	-0.285	0.003
시장점유율	0.8564	0.239	3.588	0.000	0.389	1.324
정규직:대출인력	-0.0035	0.001	-2.980	0.003	-0.006	-0.001

Omnibus:	876.695	Durbin-Watson:	0.853
Prob(Omnibus):	0.000	Jarque-Bera (JB):	4957.825
Skew:	1.552	Prob(JB):	0.00
Kurtosis:	9.161	Cond. No.	1.15e+05

Warnings:

[1] Standard Errors are heteroscedasticity robust (HC0)

[2] The condition number is large, 1.15e+05. This might indicate that there are strong multicollinearity or other numerical problems.



## 표6

```

=====
OLS Regression Results
=====
Dep. Variable:      규모      R-squared:      0.545
Model:              OLS      Adj. R-squared:    0.543
Method:             Least Squares      F-statistic:    178.9
Date:               Mon, 06 Jun 2022    Prob (F-statistic): 4.85e-239
Time:               15:07:45          Log-Likelihood:  -14799.
No. Observations:   2500             AIC:           2.962e+04
Df Residuals:       2491             BIC:           2.967e+04
Df Model:           8
Covariance Type:    HCO
=====
               coef      std err      z      P>|z|      [0.025      0.975]
-----
Intercept      19.0100      8.435      2.254      0.024      2.477      35.542
정규직          1.0748      0.134      8.049      0.000      0.813      1.337
대출인력        0.8566      0.248      3.455      0.001      0.371      1.342
산업재산권       0.2298      0.129      1.780      0.075     -0.023      0.483
기업성장단계     6.5283      2.146      3.042      0.002      2.322     10.735
부채            0.0018      0.000      3.978      0.000      0.001      0.003
지분구조       -0.1390      0.074     -1.888      0.059     -0.283      0.005
시장점유율      0.8489      0.237      3.579      0.000      0.384      1.314
정규직:대출인력 -0.0036      0.001     -3.002      0.003     -0.006     -0.001
=====
Omnibus:          878.181      Durbin-Watson:    0.853
Prob(Omnibus):    0.000      Jarque-Bera (JB): 4968.592
Skew:             1.555      Prob(JB):         0.00
Kurtosis:         9.167      Cond. No.         1.07e+05
=====

Warnings:
[1] Standard Errors are heteroscedasticity robust (HCO)
[2] The condition number is large, 1.07e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

```

## 표7

```

=====
OLS Regression Results
=====
Dep. Variable:      규모      R-squared:      0.544
Model:              OLS      Adj. R-squared:    0.542
Method:             Least Squares      F-statistic:    199.2
Date:               Mon, 06 Jun 2022    Prob (F-statistic): 4.66e-235
Time:               15:08:53          Log-Likelihood: -14802.
No. Observations:   2500             AIC:           2.962e+04
Df Residuals:       2492             BIC:           2.967e+04
Df Model:           7
Covariance Type:    HCO
=====
               coef      std err      z      P>|z|      [0.025      0.975]
-----
Intercept      19.1178      8.460      2.260      0.024      2.537      35.698
정규직          1.1087      0.132      8.402      0.000      0.850      1.367
대출인력        0.8530      0.253      3.377      0.001      0.358      1.348
기업성장단계     7.0093      2.162      3.242      0.001      2.772     11.247
부채            0.0019      0.000      4.073      0.000      0.001      0.003
지분구조       -0.1520      0.073     -2.080      0.038     -0.295     -0.009
시장점유율      0.8920      0.238      3.743      0.000      0.425      1.359
정규직:대출인력 -0.0037      0.001     -3.034      0.002     -0.006     -0.001
=====
Omnibus:          874.151      Durbin-Watson:    0.848
Prob(Omnibus):    0.000      Jarque-Bera (JB): 4882.049
Skew:             1.551      Prob(JB):         0.00
Kurtosis:         9.103      Cond. No.         1.07e+05
=====

Warnings:
[1] Standard Errors are heteroscedasticity robust (HCO)
[2] The condition number is large, 1.07e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

```