

# 1. 위의 그래프를 그리는데 사용된 단기적 사업 중요도 문항과 중장기적 사업 중요도 문항 18개 문항에 대한 응답을 사용하여 응답자를 분류하여 보시오.

## 1.1 전체 문항 응답 경향

응답자들을 집단으로 분류하기 전에, 문항에 따라 응답자들의 응답 패턴이 발견되는지 살펴보았습니다. 아래 그림은 단기적 관점 및 중장기적 관점 18개 문항에 대한 응답 간 상관관계를 연결망 그래프를 그린 것입니다. 예를 들어서, 어떤 항목에 대해서 응답을 10점 줄 때 다른 항목에 대한 응답이 10점, 9점일 때 9점 등 완벽히 일치한다면 상관관계가 1이 되고 완전히 예측불가라면 상관관계가 없는 것을 그린 것입니다. 문항간 응답에 대해 통계적으로 유의미한 상관관계를 구하고, 상관관계가 없거나 해당 상관관계가 전체 상관관계의 평균보다 낮은 경우는 관계가 없는 것으로 판단했습니다. 응답 척도상 통계적으로 우연히 나타날 수 있는 확률이라고 본 것입니다. 그런 다음, 평균 이상으로 관계가 있는 상관관계는 선을 이어서 그래프를 그렸습니다. 선의 굵기는 상관관계의 강도입니다.

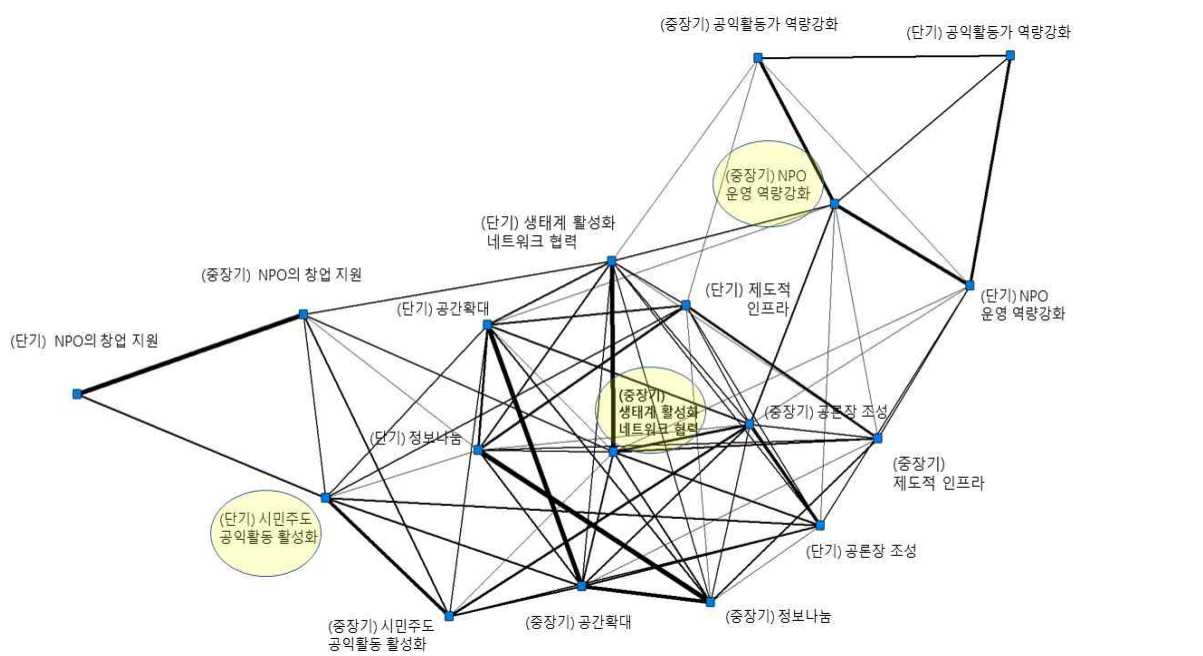


그림 1 응답 패턴에 대한 상관관계 연결망 분석 그래프

당연하게도 동일한 항목에 대한 단기 응답과 중장기 응답이 연동합니다. 그런데 해당 그래프에서 보고 싶은 것은 동일한 항목에 대한 단기 및 중장기 응답의 상관관계가 아니라, 상이한 항목에 대한 단기 및 중장기 응답이 어떻게 연관되어 있는지입니다. 해당 그래프를 보면, 응답 패턴이 크게 2가지로 정리될 수 있음을 알 수 있습니다. 우선, 창업지원과 공익활동 활성화에 대한 응답패턴은 공익활동가 역량 강화 및 운영 역

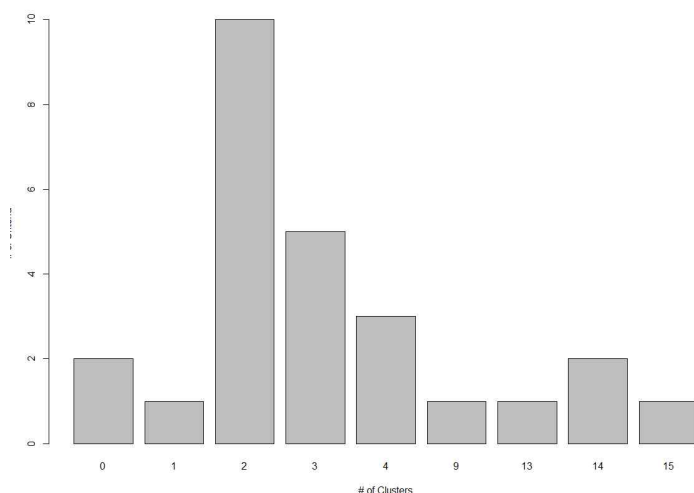
량 강화와 대한 응답패턴과 상이합니다. 따라서, 이들간에는 서로 연결된 선이 없습니다. 두 번째 패턴은 생태계 활성화 네트워크 협력, 공론장 조성, 제도적 인프라, 그리고 정보나눔 및 공간 확대에 관한 응답은 매우 밀접한 관계가 있습니다. 이를 바탕으로 관찰할 수 있는 사항을 응답자 분류 관점에서 보면, 첫째, 역량 강화에 대한 응답패턴이 창업지원과 공익활동 활성화에 대한 응답패턴과 관계가 없으므로, 이들 항목에 대한 응답이 어떻게 이루어졌느냐에 따라 응답자가 분류될 가능성이 높습니다. 둘째, 제도적 인프라, 구축, 생태계 네트워크 등의 과제에 대한 응답은 전반적으로 유사하지만, 시민 주도 공익활동 활성화, 창업 지원 등에 대한 문항과 상관관계를 더 많이 가지고 있습니다. 역량강화 응답의 오른쪽 위보다 왼쪽 아래에서부터 선이 촘촘하게 연결이 되고 있습니다. 그러므로, 다음과 같은 응답 패턴이 발견되었습니다.

1. 생태계를 활성화시키는 데 있어서 새로운 조직과 활동의 유입으로 생태계의 활력이 커져야 한다고 보는 견해 (e.g. 창업, 공익활동 활성화)
2. 현재 있는 조직과 인력의 역량 강화가 더욱 중요하다고 보는 견해. (활동가 및 운영 역량 강화)
3. 잠재적으로는 공익활동 활성화와 창업지원 등의 항목에 대해 다른 이들보다 더 높은 중요도를 부여하면서 제도적 인프라, 공론장 등 Bridging 역할을 강조하는 견해

그렇다면 이제 이러한 응답 패턴이 응답자들의 분포로 나타나는지 살펴보겠습니다.

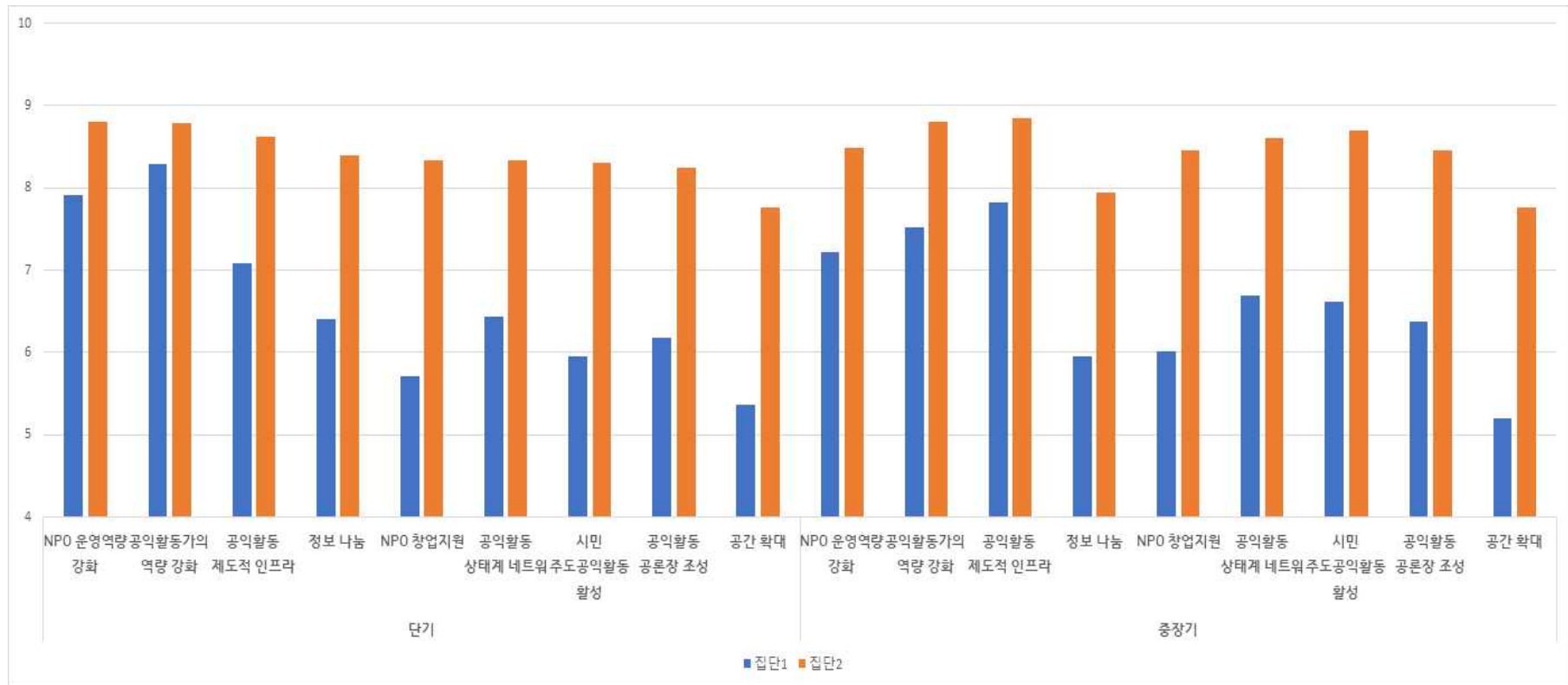
## 1.2. 응답자 분류

18개의 문항에 대한 응답을 K-means 방법을 활용, 응답자 집단을 구분하였습니다. K-means 방법이 가진 단점을 보완하고자 PAM 방법을 사용하여 검증하였으나 여전히 2개의 집단으로 구분되어 2개의 집단 수가 가장 적절한 것으로 판단되었습니다.



왼쪽의 그래프는 알고리즘이 판단한 적절한 집단의 수를 그래프로 그린 것입니다.

그렇다면 알고리즘이 2개의 집단으로 구분하였으므로, 이들 구분된 집단이 응답에 있어서 어떠한 차이가 나타나는지를 확인하여 보았습니다. 아래 그래프는 각 문항에 대한 응답 평균을 집단에 따라 구분하여 그린 것입니다.



그래프는 앞서 응답 패턴에서 확인한 사항들이 실제 응답자들의 분포 차이로 나타나고 있음을 보여주고 있습니다. 역량강화의 2가지 항목에 대해 높은 점수를 주고 제도적 인프라를 중요한 사항으로 응답한 다음 나머지 항목에 대해서는 상대적으로 저평가하고 있는 집단과 전반적으로 골고루 높은 점수를 주고 있는 집단 그 중에서도 중장기적 중요도에 있어서 생태계 네트워크, 공익활성화, 공론장, 창원지원 등이 모두 높은 점수를 준 2개의 집단이 구분되었습니다. 첫 번째 집단은 85명, 두 번째 집단은 126명으로 구분되어서 약 40%, 60%의 비율로 구분되었습니다. 첫 번째 집단을 “역량 강화 주의자”이라 명명하고 두 번째 집단을 “생태계 활성화 주의자”으로 칭하겠습니다.

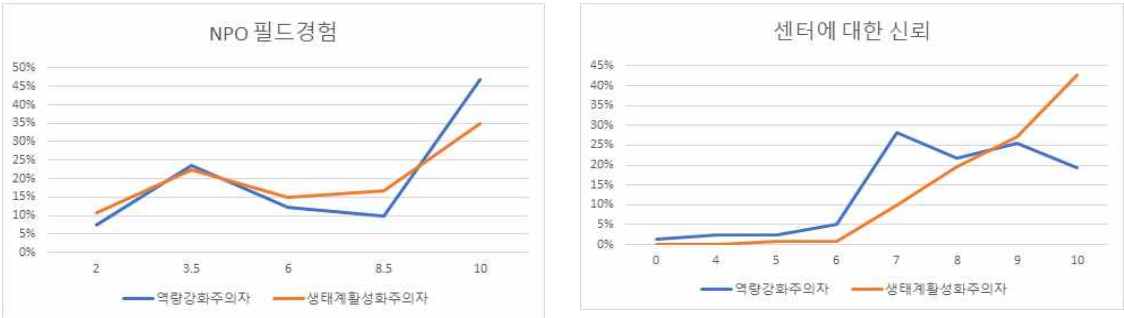
2. 1)에서 응답자들을 응답 패턴에 따라 응답자들을 분류한 프로파일링을 사업 중요도 문항이 아닌 다른 설문 문항에 대한 응답에 적용하여, 어떤 특징을 가진 응답자들이 해당 분류에 속하는지 응답자들 분류에 따른 응답자들의 성격을 특정하여 보시오.

### 2.1. 집단간 타 변수 평균 차이

역량강화 주의자와 생태계 활성화 주의자는 그럼 어떤 사람들인지 다른 변수들에 대한 통계적 차이가 있는지부터 살펴보았습니다. 기본 통계 테스트, 평균차 비교입니다.

항목	역량강화주의자	생태계활성화주의자	통계적 유의도
센터 추천한 경험	82%	89%	
미래에 추천할 가능성	7.84점	8.85점	*** p<.001
현재 정보 우선성	6.12	7.9	*** p<.001
향후정보우선성	6.89	8.33	*** p<.001
센터 신뢰성	7.98	8.98	*** p<.001
알고 지낸 기간	3.47년	3.60년	
NPO 필드 경험	7.24년	6.82년	

위의 결과는 센터를 다른 이에게 추천한 경험이나, 센터를 알고 지낸 기간 그리고 NPO 필드에 그동안 몸담은 기간에는 두 집단간 유의한 통계적 차이가 없는 것을 보여줍니다. 그러나 그보다 더 중요하게, 미래에 추천한 가능성에서의 1점이라는 차이, 그리고, 현재 NPO에 관한 정보가 필요할 때 센터를 우선적으로 찾아보는 정도, 그리고 향후에도 우선할 가능성, 그리고 무엇보다 센터에 대한 신뢰도가 차이가 있는 것을 발견했습니다. 다시 말해 2개의 집단을 구분하여 보면, 상대적으로 센터에 대한 신뢰도가 떨어지면서 센터가 보유하고 있는 정보를 찾아볼 가능성 역시 떨어지는 특히 현재 센터에 있는 정보를 우선할 가능성이 떨어지는 사람들이 역량강화주의자입니다. 반면에, 생태계 활성화 주의자들은 센터의 정보를 현재 우선하는 정도, 앞으로도 우선할 가능성, 현재 센터에 대한 신뢰 모두 상대적으로 높습니다. 아래 그래프는 통계적 차이가 있을 경우 없을 경우에 대한 직관적 비교를 위해 분포를 비교한 것입니다.



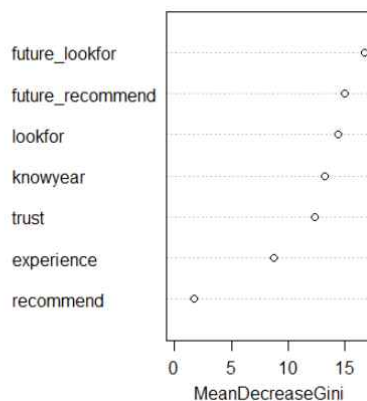
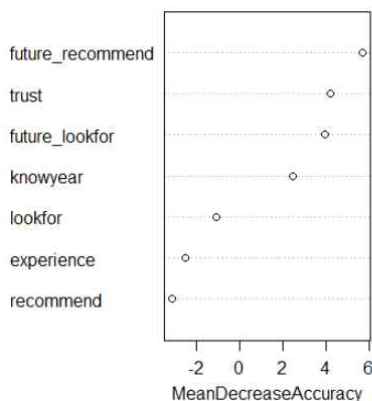
## 2.2. 응답집단 구분에 영향을 미치는 변수 중요도 확인

이제 역으로 질문의 프레임을 바꾸어 보겠습니다. 센터에 대한 의견차이가 사업의 중요도에 대한 견해 차이로 나타나는지 살펴보기로 하였습니다. 응답집단 구분에 영향을 미치는 변수를 통해서 앞서 K-means 분류에 의한 집단을 예측 할 수 있는지 생각해 보는 것입니다. 이는 센터의 사업 중요도에 대한 견해 차이가 단순히 NPO 필드에 대한 견해 차이라기보다는 센터의 역할에 대한 견해차, 센터에 대해 본인들이 가지고 있는 태도 등에 의해서 나타날 것이라는 예상을 하는 것입니다.

만약 예측의 성능이 매우 낮다면, 일차적으로 센터에 대한 의견차이가 센터의 사업 중요도 판단에 주요한 원인이라기보다는 NPO 필드에 대한 구조적 판단 차이로 인해 센터의 사업 중요도에 대한 차이가 나타났다고 생각할 가능성이 생깁니다. 물론, 기계학습 알고리즘의 종류에 따라 예측정도가 차이가 날 수 있지만, 일반적으로 볼 때 현재 숫자 정도의 데이터 포인트에 대한 기계학습의 성능은 뛰어난 편입니다. 따라서, 센터에 대한 견해 변수가 효용성이 없다고 볼 여지가 생긴다는 것입니다. 이와 반대로 센터에 대한 의견 변수들로 사업 중요도에 대한 견해가 거의 완벽하게 예측이 된다면, 적어도 센터에 대한 자신의 의견이 앞으로 센터가 나아가야 할 방향에 대한 견해에 상당부분 영향을 미쳤다고 할 수 있을 것입니다.

여러 알고리즘을 동원할 수 있습니다만, 여기서는 랜덤포레스트를 활용하여 보았습니다. 그 결과 98%의 예측 정확성을 보였으며 오류율은 2%에 불과하였습니다.

		랜덤포레스트 예측	
		역량강화	생태계활성화
K-means 분류	역량강화	61	2
	생태계활성화	2	134

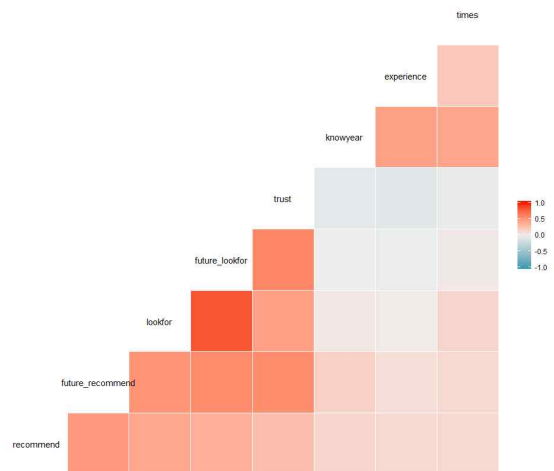


예측하는데 중요한 변수는 앞서 평균차 비교에서 통계적으로 유의미한 차이를 보이는 변수들이었는데, 주로 향후 추천할 가능성, 센터에 대한 신뢰 그리고 향후 정보 우선성

을 센터에 두는지에 대한 여부 등의 순서였으며, 분류가 잘 이루어지도록 하는데 있어서 중요 변수는 이와 비슷한 순서를 보이는 가운데 현재 정보 우선성이 포함되는 것으로 나타났습니다.

확인을 위해 변수들간 상관관계도 한번 살펴보았습니다. 색깔이 진해질수록 상관관계가 높은 것인데, 신뢰의 경우는 다른 변수가 상관관계가 그렇게 높다고 보기는 어려운 것으로 확인되었습니다.

앞선 기계학습 결과를 시험해 보기 위해, 다음의 값을 가지는 경우를 상정해서 해당 분류기에 시험해 보았습니다. 어떤 사람의 경우 센터에 대한 신뢰도가 7점이고, 향후 추천할 가능성도 7점,



현재 정보 우선성은 8점 향후에도 우선할 가능성 8점에 센터를 알게된지 3.5년, NPO 필드 경험 6년인 경우, 분류기는 해당 응답자를 “역량 강화주의자”로 분류하였습니다. 이 사람의 프로파일에서 다른 항목들에 대한 점수를 그대로 두고, 센터에 대한 신뢰도를 9점으로 올려도 여전히 역량강화주의자로 분류되었고, 마찬가지로, 미래에 정보를 우선할 가능성만을 9점으로 올려도 여전히 역량강화 주의자로 분류되었습니다. 심지어, 신뢰도가 9점, 현재 정보우선성과 향후 우선성 모두 9점으로 하여도 역량강화주의자가 되었습니다. 오로지 향후에 추천가능성과 나머지 점수들을 모두 동시적으로 높은 점수로 올렸을 때에만, 생태계활성화주의자로 분류되었습니다. 즉, 센터를 향후 타인에게 추천할 가능성이 “매우” 높으면서 정보우선성에 대한 가치도 높고 따라서 센터에 대한 신뢰도도 높아야 생태계활성화주의자로 분류될 수 있었다는 이야기입니다. 따라서, 센터를 어떻게 생각하느냐가 센터의 사업 방향성에 대한, 단기, 중장기 사업 중요도에 관한 판단에 영향을 미쳤다고 볼 수 있습니다. 따라서 센터에 대한 태도와 응답간 관계를 놓고 서술하면 아래와 같습니다.

첫째, 센터에 대한 신뢰가 상대적으로 높으면서 정보 활용에도 매우 적극적인 사람들은 센터의 역할이 역량강화가 아니라 조직들을 엮어내는 중간조직으로서 Bridging을 하며 생태계를 더욱 활성화시키면서 제도적 인프라를 구축하는 등 전체 판을 키워야 한다고 보고 있습니다.

반면에 이와 견해가 다른 사람들은 둘째, 센터의 정보를 현재 이용하고 향후에도 필요하면 이용하긴 하겠으나, 센터에 대한 신뢰는 상대적으로 적고 센터에 대한 활용도 역시 적은 사람들은, 센터가 해야 할 일이 전체 판을 키우는 것이 아니라 NPO 조직과 개인 활동가들의 역량을 키우는 일에 더 높은 순위를 두어야 한다고 봅니다.

### 3. 해당 결과를 가지고 인사이트를 도출하시오

본 조사 결과는 앞으로의 조직의 활동 방향성에 관해 매우 다양한 이야기를 보여주고 있습니다. 무엇보다 조직의 지속가능성(Substantiality)과 앞으로의 사업 계획에 관한 기록에 서 있는 것이 아닌가 하는 고찰을 주고 있습니다. 분석 결과, 응답집단에 상관없이 어느 정도 응답자들간 공감하고 사항은, 센터에 대한 신뢰나 센터를 타인에게 추천할 가능성이 - 집단마다 차이가 있음에도 불구하고 - 높은 편이라는 점입니다. 역량 강화 집단만 보더라도 추천한 경험이 82%였고, 신뢰도도 8점여 육박하였습니다. 상대적으로 타 집단와의 차이는 있고, 방향성에 대한 견해는 다르겠지만, 센터가 가지는 역할 그 자체에 대한 컨센서스는 어느 정도 있다고 보아도 무방하겠습니다. 다만, 다음과 같은 사항은 앞으로 더 고려해야 할 것 같습니다.

1. 역량 강화주의자들이 상대적으로 정보 우선성에 대한 평가가 떨어지는 이유는 ‘정보의 종류’에서 나타나는 것이 아닌가 생각해 볼 필요가 있습니다. 앞으로 다양한 이해관계자들을 만나면서 센터가 제공하는 정보에 대한 평가를 더욱 자세히 들여다 볼 필요가 있을 것으로 생각됩니다.

2. 장기적으로, 특히 지속가능성의 관점에서 볼 때, 역량강화와 생태계 활성화는 모두 필요한 사업인데 사업의 우선성을 두고 차이가 나타나는 건, 본고가 밝힌 센터에 대한 태도나 의견 차이일수도 있지만, 현재 전체 NPO 필드의 ‘성숙도’에 대한 견해 차이에서 나타나는 것일 수도 있습니다. 이러한 ‘성숙도’ 판단이 사업 중요도에 어떤 방식으로 영향을 미치는지, 다각도로 검토해야 할 필요가 있습니다. 센터에 대한 무한 신뢰를 보낸다 하여도, 시장이 성숙하지 않았다 생각하면 견해가 달라지는 것은 당연하게 보입니다.



3. 현재 인력난이라든가 현 NPO 조직들이 겪는 어려움을 조금 더 자세히 들여다 볼 필요성도 있습니다. 맨 처음 문제 설명에서 제시된 설문 결과 - 사분면에 표시된 응답 그래프 -를 보면 중요하다고 보고는 있으나 창업지원과 같은 신생 조직 지원, 제도적 인프라 구축, 생태계 네트워크 협력 등 조직과 조직의 연결보다, 지금 현존 조직의 운영 역량 강화라든가, 활동가 역량 강화에 대해 집단과 상관없이 매우 단기적 시급성을 매우 높게, 중장기적 시급성은 활동가 역량의 경우는 제도적 인프라 구축 다음으로 높게 평가한 것을 알 수 있습니다. 전체 필드에 대한 평가와 더불어서, 응답자들의 조직 상황이 불안정하다는 인상을 지을 수가 없습니다. 해당 지점에 대한 고찰이 필요합니다.

본 답안 예제는 분석 결과에 대한 리포팅이 빠져있어 10쪽을 미쳐 다 채우지 않고 있습니다. 여러분이 제출한 답안에서 보이듯 분석 과정 내용에 대한 서술이 이루어지고, Cross-Validation 이라든가, MISSING 처리 등 분석 기법에 대한 설명이 추가될 경우 10장을 채우게 됩니다. 본 답안은 하나의 예시로 본 답안이 보여주고 있는 분석 방법과 그에 따른 분석 결과가 본인의 분석결과가 일치하지 않는다는 점이 평가의 대상이 되지 않습니다. 다만 본인 분석의 내적 정합성은 필요합니다. 또 한가지, 제시된 문제 자체가 데이터를 깊게 파고 들어가는 자세한 분석을 요구하고 있지는 않습니다. 본 과제는 수업시간에 배운 사항을 제대로 잘 구현했는가가 중심이며, 해당 사항이 잘 정리되었는 것입니다. 중간고사 채점 점수는 제출하신 파일에 대한 feedback과 함께 돌려드리겠습니다.

중간고사 수고하셨습니다.