

[R통계] 다중회귀분석

조성우 201823869

Department of E-business

2020-11-23

- 회귀분석
- 1. 개요
 - 연구과제
 - 가설설정
 - 데이터
 - 연구법(분석기법)
- 2. 본문
 - 2.1 데이터 준비
 - 2.1.1 데이터 불러오기
 - 2.1.2 데이터 전처리
 - 2.4 다중회귀분석
 - 2.5 모델 선택
- 3. 결론
 - 개선모델 결과해석

회귀분석

하나 혹은 그 이상의 독립변수들이 종속변수에 미치는 영향을 추정할 수 있는 통계기법으로 변수들 사이의 인과관계를 밝히고 모형을 적합하여 관심있는 변수를 예측하거나 추론하기 위한 분석방법이다.

1. 개요

연구과제

국가별 혁신점수 를 종속변수로 하여, 독립변수인 소속대륙 과 국가별 문화적 차원점수(6개) 가 종속변수에 미치는 영향을 알아보고자 한다.

이를 위해 R을 사용한 다중회귀분석 및 최적 회귀방정식을 구하기 위한 일련의 분석과정을 실시해보

고, 그 과정 및 결과를 해당 보고서에 기록하고자 한다.

가설설정

연구 가설 : 독립변수가 종속변수에 유의한 영향을 미칠 것이다.

연구 가설1 : 국가가 소속한 대륙 이 *Global Innovation Score*에 유의한 영향을 미칠 것이다.

연구 가설2 : 국가의 *Hofstede 6dimensions score* 가 *Global Innovation Score*에 유의한 영향을 미칠 것이다.

- 2-1 : 권력거리 지수(*pdi*) 가 *Global Innovation Score*에 유의한 영향을 미칠 것이다
- 2-2 : 단체귀속정도(*idv*) 가 *Global Innovation Score*에 유의한 영향을 미칠 것이다
- 2-3 : 남성주의도(*mas*) 가 *Global Innovation Score*에 유의한 영향을 미칠 것이다
- 2-4 : 불확실성 회피지수(*uai*) 가 *Global Innovation Score*에 유의한 영향을 미칠 것이다
- 2-5 : 장기지향성(*ltowvs*) 이 *Global Innvation Score*에 유의한 영향을 미칠 것이다
- 2-6 : 쾌락지향성(*ivr*) 이 *Global Innovation Score*에 유의한 영향을 미칠 것이다

데이터

데이터셋

- **global innovation index 2020** :
<https://www.globalinnovationindex.org/analysis-indicator> 에 첨부된 CSV파일(2020.nov) 을 사용
- **Hofstede 국가별 문화적 차원점수** :
<https://geerthofstede.com/research-and-vsm/dimension-data-matrix/> 에 첨부된 6-dimensions-for-website.csv 데이터를 사용

주요변수 및 특징

두 데이터셋에서 활용할 주요변수 몇가지를 요약/기술하는 항목

(1) *global innovation index*

- **country** : 비연속척도이며 Hofstede 데이터셋과 병합하기 위한 key변수이다.
- **score** (종속변수) : 해당 국가의 혁신점수로 0~100의 실수형 range를 갖는다.

(2) *Hofstede 6-dimension*

- **country** : 비연속척도이며 GII 데이터셋과 병합하기 위한 key변수이다.
- **권력거리지수(*pdi*)** : 해당 권력거리 점수가 낮을수록 보다 수평적이고 민주적인 문화이다. (독립변수)
- **단체귀속정도(*idv*)** : 개인들이 단체에 통합되는 정도로, 점수가 낮을수록 개인주의적 문화이다. (독립변수)

- 남성주의도 (*mas*) : *mas*가 높을수록, 남성주의적 특성이 짙은 사회이다. (독립변수)
- 불확실성 회피지수 (*uai*) : 불확실성에 대한 사회적 저항력으로, *uai*가 낮을수록 비체계적인 상황이나 가변적 환경을 편안히 받아들인다. (독립변수)
- 장기지향성 (*ltowvs*) : 미래와 과거중 어떤것에 중요성을 부여하는가에 관한 지표이다. (독립변수)
해당 점수가 높을수록 미래에 더 많은 중요성을 부여하며 보상을 지향하는 실용적인 문화를 조성하며, 반대로 낮을수록 과거에 더 많은 중요성을 부여하며 전통과 같은 과거와 현재에 관련한 가치를 준수하는 경향이 있다.
- 쾌락지향성 (*ivr*) : 욕구와 충동에 따라 행동하는 경향을 나타내는 지표이다.
이 점수가 낮은 사회일수록 비관적이고 냉소적인 경향을 나타낸다. 반대로 이 점수가 높은 사회는 낙관적이고 삶을 즐기려는 경향을 보인다. (독립변수)

(3) 추가변수 및 특징

- 대륙 : 독립변수로 사용하기 위해 추가해야하며, 범주변수이기 때문에 더미변수로 변환하고자 한다.
- *6 dimensions' score* : 음/양의 방향이 통일되지 않아, 필요에 따라 정확히 재부호화 시킬 필요가 있을 수 있다.

연구법(분석기법)

- 1. 데이터 준비
- 2. 다중회귀분석
- 3. 다중 회귀모델 최적화

2. 본문

2.1 데이터 준비

회귀분석을 시행하기에 앞서 Raw data를 불러오고 이에 적절하게 처리한다.

2.1.1 데이터 불러오기

```
gii_df = read.csv('GII_2020nov_main.csv', header = T)
hof_df = read.csv('Hofstede_6dimensions.csv', header = T)
```

```
knitr::kable(head(gii_df,5))
```

rank	Economy	Income.Group.Strength.Weakness.	Strength...Weakness	Score
83	Albania			27.1
121	Algeria	Weakness		19.5
80	Argentina			28.3
61	Armenia			32.6
23	Australia	Weakness		48.4

Global Innovation Index 2020 원데이터의 모습

```
knitr::kable(head(hof_df,5))
```

continent	country	pdi	idv	mas	uai	ltowvs	ivr	X	X.1	X.2
AFE	Africa East	64	27	41	52	32	40	NA	NA	NA
AFW	Africa West	77	20	46	54	9	78	NA	NA	NA
ALB	Albania	#NULL!	#NULL!	#NULL!	#NULL!	61	15	NA	NA	NA
ALG	Algeria	#NULL!	#NULL!	#NULL!	#NULL!	26	32	NA	NA	NA
AND	Andorra	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	65	NA	NA	NA

Hofstede 6dimension 원데이터의 모습

2.1.2 데이터 전처리

1. gii_df 데이터 전처리

1. 주요변수 재구성

```
attach(gii_df)
gii_df = data.frame(country = Economy, score = Score)
detach(gii_df)
```

2. hof_df 데이터 전처리

1. 국가코드 변수명 수정

```
colnames(hof_df)[1] = 'alpha3'
```

2. numeric형으로 변환하기

```
attach(hof_df)

hof_df$pdv = as.numeric(pdv)
hof_df$idv = as.numeric(idv)
hof_df$mas = as.numeric(mas)
hof_df$uai = as.numeric(uai)
hof_df$ltowvs = as.numeric(ltowvs)
hof_df$ivr = as.numeric(ivr)
detach(hof_df)
```

3. 두 데이터셋 병합

```
main_df = merge(gii_df, hof_df, by='country')

main_df = main_df[1:9] #불필요 변수 제거
```

4. 대륙변수 추가 및 할당

```
#(1) 국가별 코드(2자리/3자리) csv 불러오기
CCI = read.csv('1.CountryCodeIndex.csv', header = T, encoding='UTF-8')

#(2) 대륙별 포함국가목록(코드포함) csv 불러오기
cont_country = read.csv('2.대륙별_포함국가목록_by_alpha2.csv', header=T)

#(3) country_alpha2를 key로 Inner_Join
cont_country_df = merge(CCI, cont_country, key='country_alpha2')

#(4) cont_country_df 주요변수로 재구성
attach(cont_country_df)
cont_country_df = data.frame(alpha2 = country_alpha2, alpha3 = country_alpha3, continent=cont)
detach(cont_country_df)

#(5) alpha3를 key로 main_df와 cont_count_df를 Inner_Join
cont_country_df = merge(main_df, cont_country_df, key='alpha3') #대륙변수가 할당됨

cont_country_df$continent[cont_country_df$continent == '중동'] = '아시아' # 중동을 아시아로 할
cont_country_df$continent[cont_country_df$continent == '중남미'] = '남미'
```

```
#(6) 할당된 대륙변수를 factor로 변환
cont_country_df$continent[cont_country_df$continent == '중남미'] = '남미'
cont_country_df$continent[cont_country_df$continent == '중동'] = '아시아'
cont_country_df$continent = factor(cont_country_df$continent)
main_df = main_df[, -1]

#(7) 불필요해진 국가코드 제외
main_df = cont_country_df[, -c(1, 10)]
```

5. 대륙변수를 더미변수로 변환 [fastDummies 패키지]

```
library(fastDummies)

#더미변수 추가
main_df = dummy_cols(.data = main_df, select_columns = 'continent', remove_first_dummy = FALSE)
```

```
#<<<전처리가 완료된 데이터의 모습>>>
head(main_df, 5)
```

```
##      country score pdi idv mas uai ltowvs ivr  continent continent_남미
## 1  Albania  27.1  NA  NA  NA  NA    61  15    유럽              0
## 2  Algeria  19.5  NA  NA  NA  NA    26  32    아시아              0
## 3 Argentina 28.3  49  46  56  86    20  62    남미                1
## 4  Armenia  32.6  NA  NA  NA  NA    61  NA    아시아              0
## 5 Australia 48.4  38  90  61  51    21  71 오세아니아              0
##      continent_북미 continent_아시아 continent_아프리카 continent_오세아니아
## 1              0              0              0              0
## 2              0              1              0              0
## 3              0              0              0              0
## 4              0              1              0              0
## 5              0              0              0              1
##      continent_유럽
## 1              1
## 2              0
## 3              0
## 4              0
## 5              0
```

```
str(main_df)
```

```
## 'data.frame':   91 obs. of  15 variables:
## $ country      : chr  "Albania" "Algeria" "Argentina" "Armenia" ...
```

```
## $ score : num 27.1 19.5 28.3 32.6 48.4 50.1 27.2 20.4 49.1 31.3 ...
## $ pdi : num NA NA 49 NA 38 11 NA 80 65 NA ...
## $ idv : num NA NA 46 NA 90 55 NA 20 75 NA ...
## $ mas : num NA NA 56 NA 61 79 NA 55 54 NA ...
## $ uai : num NA NA 86 NA 51 70 NA 60 94 NA ...
## $ ltowvs : num 61 26 20 61 21 60 61 47 82 81 ...
## $ ivr : num 15 32 62 NA 71 63 22 20 57 15 ...
## $ continent : Factor w/ 6 levels "남미","북미",...: 6 3 1 3 5 6 3 3 6 6 ...
## $ continent_남미 : int 0 0 1 0 0 0 0 0 0 0 ...
## $ continent_북미 : int 0 0 0 0 0 0 0 0 0 0 ...
## $ continent_아시아 : int 0 1 0 1 0 0 1 1 0 0 ...
## $ continent_아프리카 : int 0 0 0 0 0 0 0 0 0 0 ...
## $ continent_오세아니아 : int 0 0 0 0 1 0 0 0 0 0 ...
## $ continent_유럽 : int 1 0 0 0 0 1 0 0 1 1 ...
```

전처리가 완료된 최종 데이터의 모습과 구조

2.4 다중회귀분석

```
#>다중회귀분석
reg = lm(score ~ pdi + idv + mas + uai + ltowvs + ivr + continent_북미 + continent_아시아 +
summary(reg)
```

```
##
## Call:
## lm(formula = score ~ pdi + idv + mas + uai + ltowvs + ivr + continent_북미 +
## continent_아시아 + continent_아프리카 + continent_오세아니아 +
## continent_유럽, data = main_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.9567  -3.2336   0.0881   3.0031  14.9518
##
## Coefficients: (2 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.69154    7.85473   2.252  0.02901 *
## pdi           -0.09035    0.05432  -1.663  0.10291
## idv            0.08518    0.06515   1.307  0.19746
## mas           -0.01885    0.04246  -0.444  0.65921
## uai           -0.08558    0.04442  -1.927  0.06007 .
## ltowvs         0.24332    0.04660   5.222 3.96e-06 ***
## ivr            0.24877    0.05451   4.564 3.62e-05 ***
## continent_북미 10.72945    7.07916   1.516  0.13631
## continent_아시아 10.05569    3.81479   2.636  0.01133 *
## continent_아프리카 NA         NA         NA      NA
## continent_오세아니아 6.19143    5.69140   1.088  0.28220
## continent_유럽 11.30466    3.53949   3.194  0.00251 **
```

```
## continent_남미      NA      NA      NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.042 on 47 degrees of freedom
## (33 observations deleted due to missingness)
## Multiple R-squared:  0.758, Adjusted R-squared:  0.7065
## F-statistic: 14.72 on 10 and 47 DF, p-value: 2.188e-11
```

```
#]결과해석
```

표준화 계수 출력

```
#>>표준화계수출력
library(QuantPsyc)
```

```
## Loading required package: boot
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'QuantPsyc'
```

```
## The following object is masked from 'package:base':
##
##      norm
```

```
lm.beta(reg)
```

```
##           pdi           idv           mas
##      -0.170096131      0.175367086     -0.034861124
##           uai           ltowvs           ivr
##      -0.182037696      0.482729758      0.485025116
## continent_북미 continent_아시아 continent_아프리카
##      0.126335209      0.398291712      0.000000000
## continent_오세아니아 continent_유럽 continent_남미
##      0.186584653     -0.004076763      0.002789735
```

다중공산성 진단


```
#>> 다중공선성 진단을 위한 VIF값 구하기
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:boot':
##
##      logit
```

```
vif(reg) # 에러 : there are aliased coefficients in the model 완전공분산성이 존재한다.
```

```
## Error in vif.default(reg): there are aliased coefficients in the model
```

에러메세지로 완전 공분산성이 존재함을 확인

완전공분산성을 제거하고 회귀모델 재설계

```
#>>> 완전공분산성을 제거하기 위해 결과로 NA를 가진 두 대륙을 제외하고 회귀모델을 다시 설계한다.
reg1 = lm(score ~ pdi + idv + mas + uai + ltowvs + ivr + continent_북미 + continent_아시아 +
summary(reg1)
```

```
##
## Call:
## lm(formula = score ~ pdi + idv + mas + uai + ltowvs + ivr + continent_북미 +
##      continent_아시아 + continent_오세아니아 + continent_유럽,
##      data = main_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.9567  -3.2336   0.0881   3.0031  14.9518
##
## Coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.69154     7.85473   2.252  0.02901 *
## pdi             -0.09035     0.05432  -1.663  0.10291
## idv              0.08518     0.06515   1.307  0.19746
## mas             -0.01885     0.04246  -0.444  0.65921
## uai             -0.08558     0.04442  -1.927  0.06007 .
## ltowvs          0.24332     0.04660   5.222 3.96e-06 ***
## ivr             0.24877     0.05451   4.564 3.62e-05 ***
```

```
## continent_북미      10.72945    7.07916    1.516    0.13631
## continent_아시아    10.05569    3.81479    2.636    0.01133 *
## continent_오세아니아  6.19143    5.69140    1.088    0.28220
## continent_유럽      11.30466    3.53949    3.194    0.00251 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.042 on 47 degrees of freedom
## (33 observations deleted due to missingness)
## Multiple R-squared:  0.758, Adjusted R-squared:  0.7065
## F-statistic: 14.72 on 10 and 47 DF, p-value: 2.188e-11
```

#>>> #완전공분산성이 해결되어 *vif*의 결과가 출력된다.

vif(reg1) #결과해석 : 모든 *vif*값이 2 미만으로 나타나고 있어 다중공선성으로 인한 문제는 없는것으로

```
##                pdi                idv                mas
##          2.030990          3.494393          1.198183
##                uai                ltowvs                ivr
##          1.733581          1.659749          2.193711
## continent_북미 continent_아시아 continent_오세아니아
##          1.349304          4.433764          1.713670
## continent_유럽
##          4.953127
```

개선모델 다중회귀 결과표 및 도표 출력

```
#>> 'sjPlot' 패키지를 이용한 결과표 및 도표 출력
# 다중회귀분석 결과표 작성
library(sjPlot)
```

```
## Registered S3 methods overwritten by 'lme4':
##   method                      from
##   cooks.distance.influence.merMod car
##   influence.merMod              car
##   dfbeta.influence.merMod       car
##   dfbetas.influence.merMod      car
```

```
## Install package "strengjacke" from GitHub (`devtools::install_github("strengjacke/stre
```

```
tab_model(reg1, show.se = T, show.ci = F, show.stat= T)
```

<i>Predictors</i>	<i>Estimates</i>	<i>std. Error</i>	<i>Statistic</i>	<i>p</i>
(Intercept)	17.69	7.85	2.25	0.029
pdi	-0.09	0.05	-1.66	0.103
idv	0.09	0.07	1.31	0.197
mas	-0.02	0.04	-0.44	0.659
uai	-0.09	0.04	-1.93	0.060
ltowvs	0.24	0.05	5.22	<0.001
ivr	0.25	0.05	4.56	<0.001
continent_遺곳	10.73	7.08	1.52	0.136
continent_북뎡북	10.06	3.81	2.64	0.011
continent_뽕꺼북뎡북	6.19	5.69	1.09	0.282
continent_쑤옴	11.30	3.54	3.19	0.003
Observations	58			
R ² / R ² adjusted	0.758 / 0.706			

한글을 사용

```
tab_model(reg1, show.se = T, show.ci = F, show.stat= T,
  pred.labels = c("(Intercept)", "pdi", "idv", "mas","uai",
    "ltowvs", "ivr", 'Dummy_NorthAmerica', 'Dummy_ASIA', 'Dummy_Oceania', '
  file ="multiple_regression.html")
```

Global Innovation Index

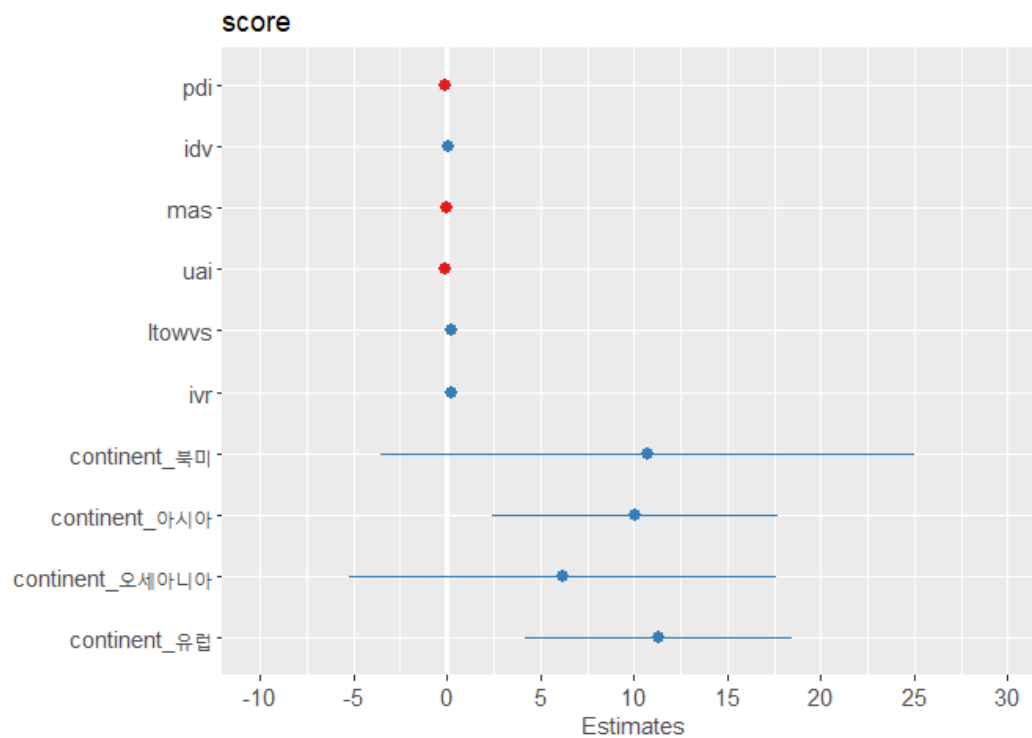
<i>Predictors</i>	<i>Estimates</i>	<i>std. Error</i>	<i>Statistic</i>	<i>p</i>
(Intercept)	17.69	7.85	2.25	0.029
pdi	-0.09	0.05	-1.66	0.103
idv	0.09	0.07	1.31	0.197
mas	-0.02	0.04	-0.44	0.659
uai	-0.09	0.04	-1.93	0.060
ltowvs	0.24	0.05	5.22	<0.001
ivr	0.25	0.05	4.56	<0.001
Dummy_NorthAmerica	10.73	7.08	1.52	0.136
Dummy_ASIA	10.06	3.81	2.64	0.011
Dummy_Oceania	6.19	5.69	1.09	0.282

Dummy_Europe	11.30	3.54	3.19	0.003
Observations	58			
R ² / R ² adjusted	0.758 / 0.706			

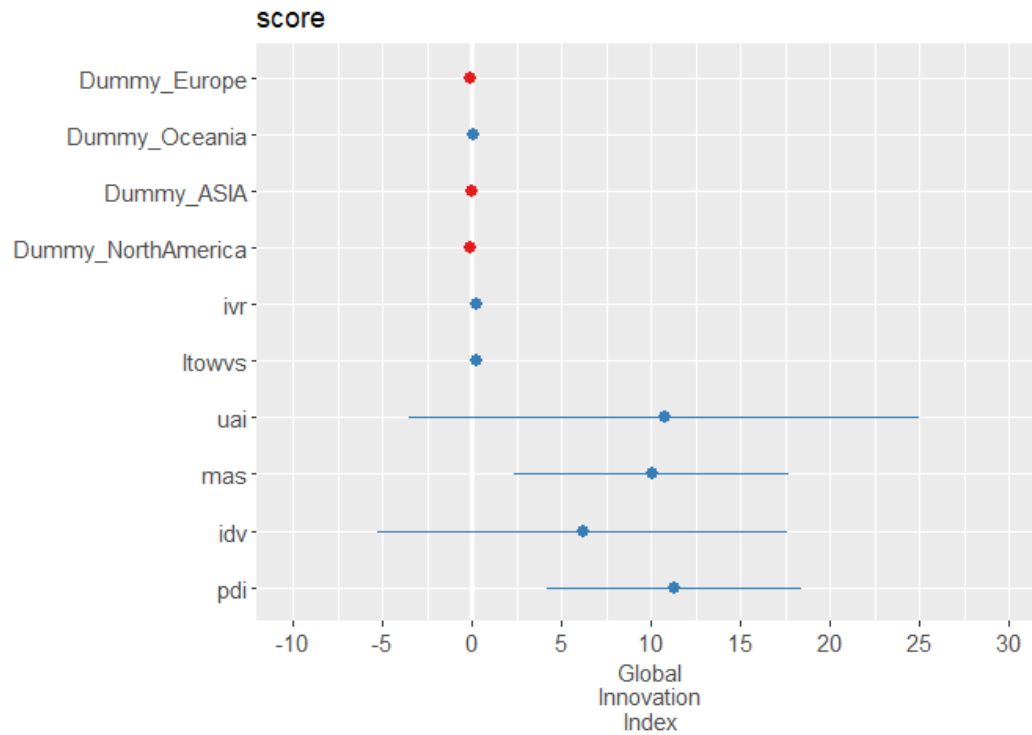
```
file.show("multiple_regression.html")
```

다중회귀분석 도표 작성

```
#>> 'sjPlot' 패키지를 이용한 다중회귀분석 도표 작성
set_theme(axis.title.size = 1.0, axis.textsize = 1.0)
plot_model(reg1, type = "est", wrap.labels=10)
```

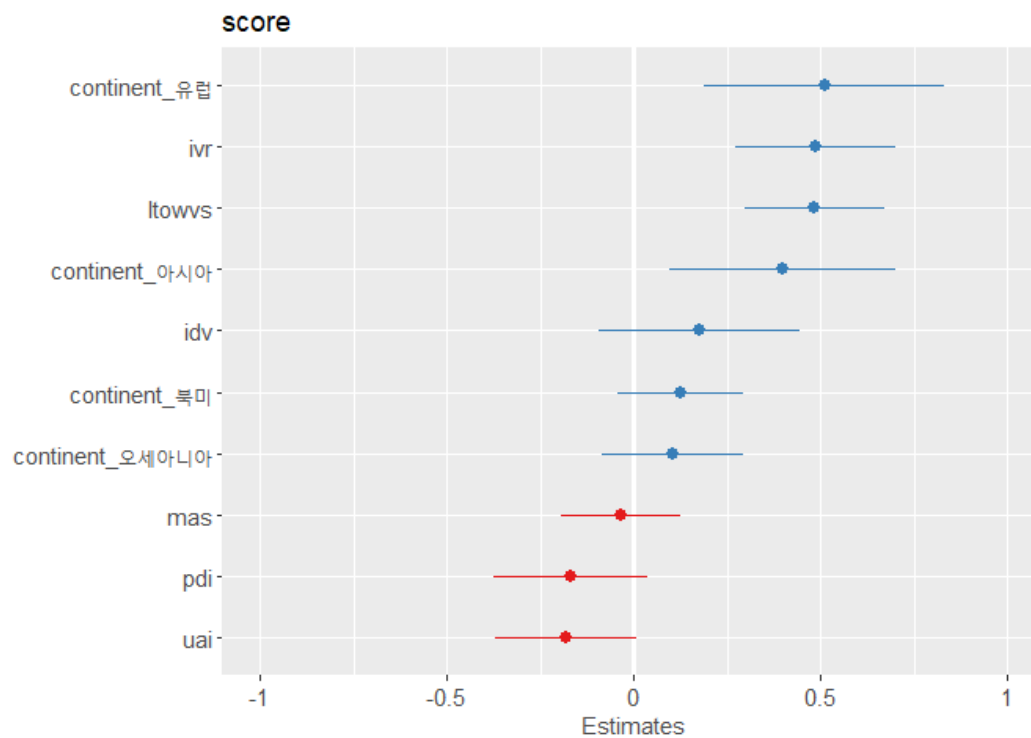


```
plot_model(reg1, type = "est",
  axis.labels=c("pdi", "idv", "mas", "uai",
    "ltowvs", "ivr", 'Dummy_NorthAmerica', 'Dummy_ASIA', 'Dummy_Oceania', '
  axis.title="Global Innovation Index", wrap.labels=5)
```

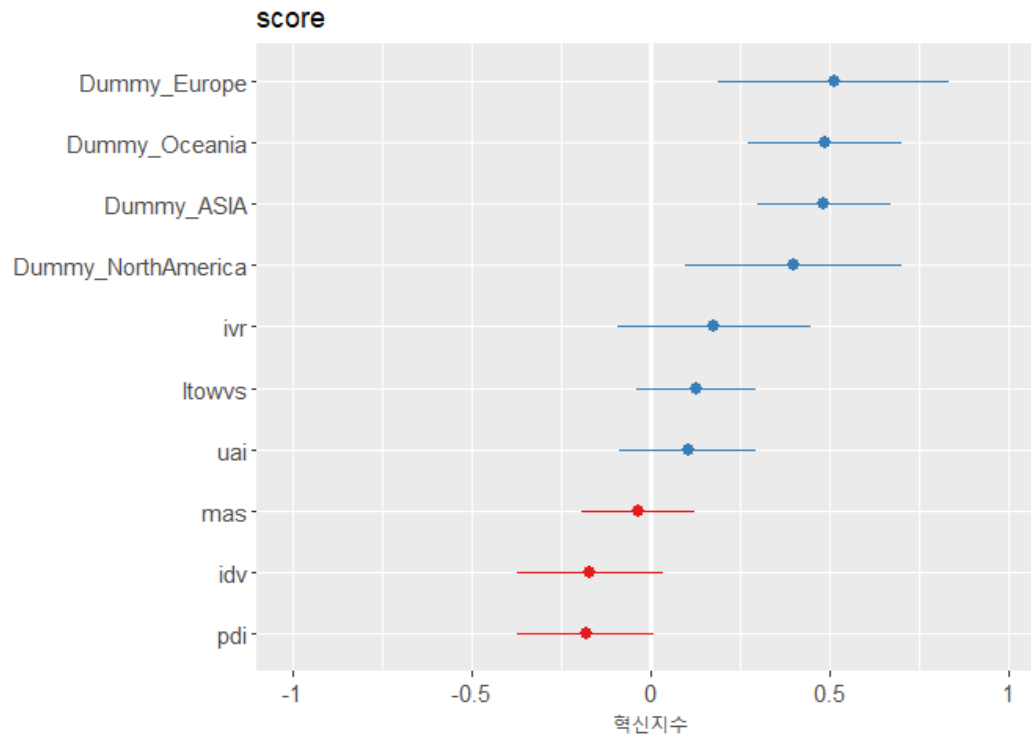


다중회귀분석 도표에 표준화계수를 이용해 작성

```
#>>> 'sjPlot' 패키지를 이용한 다중회귀분석 도표에 '표준화 계수'를 이용하여 작성
plot_model(reg1, type = "std", sort.est = T, wrap.labels=5)
```



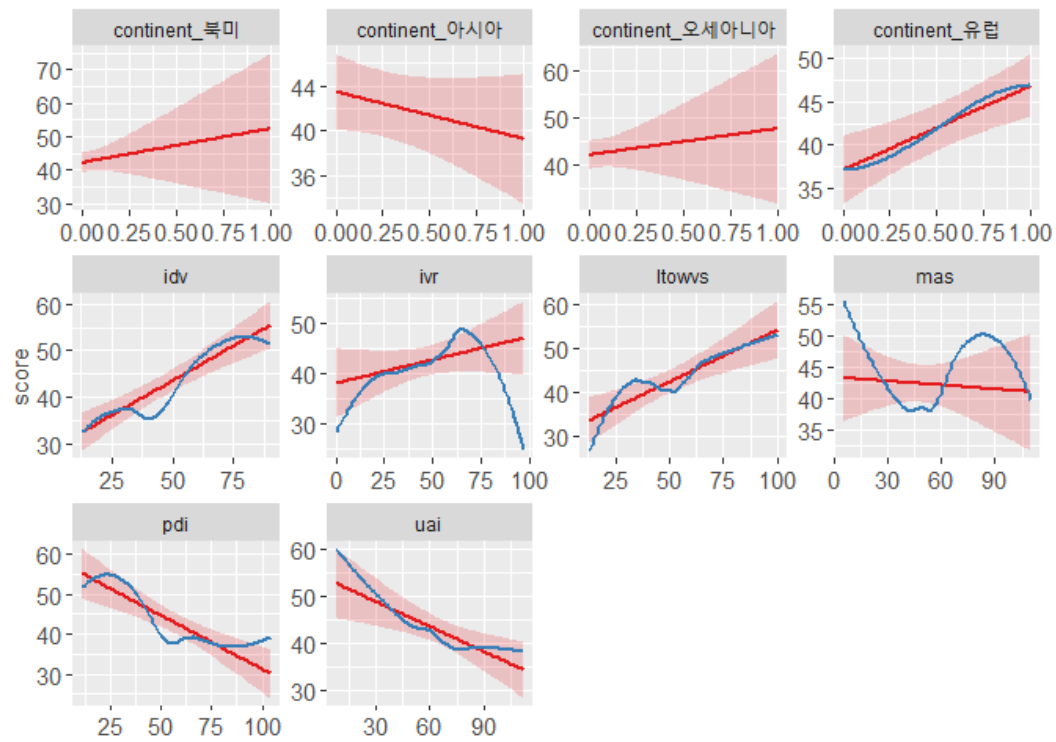
```
plot_model(reg1, type = "std", sort.est = T,
            axis.labels=c("pdi", "idv", "mas", "uai",
                          "ltowvs", "ivr", 'Dummy_NorthAmerica', 'Dummy_ASIA', 'Dummy_Oceania', '
            axis.title="혁신지수", wrap.labels=5)
```



독립변수별 다중회귀 *Slope* 그래프 시각화

```
#>>> slope 시각화를 사용한 다중회귀 선형 그래프 시각화
plot_model(reg1, y=, type="slope") # 양과 음의 관계 및 오차 정도를 변수별로 확인 가능
```

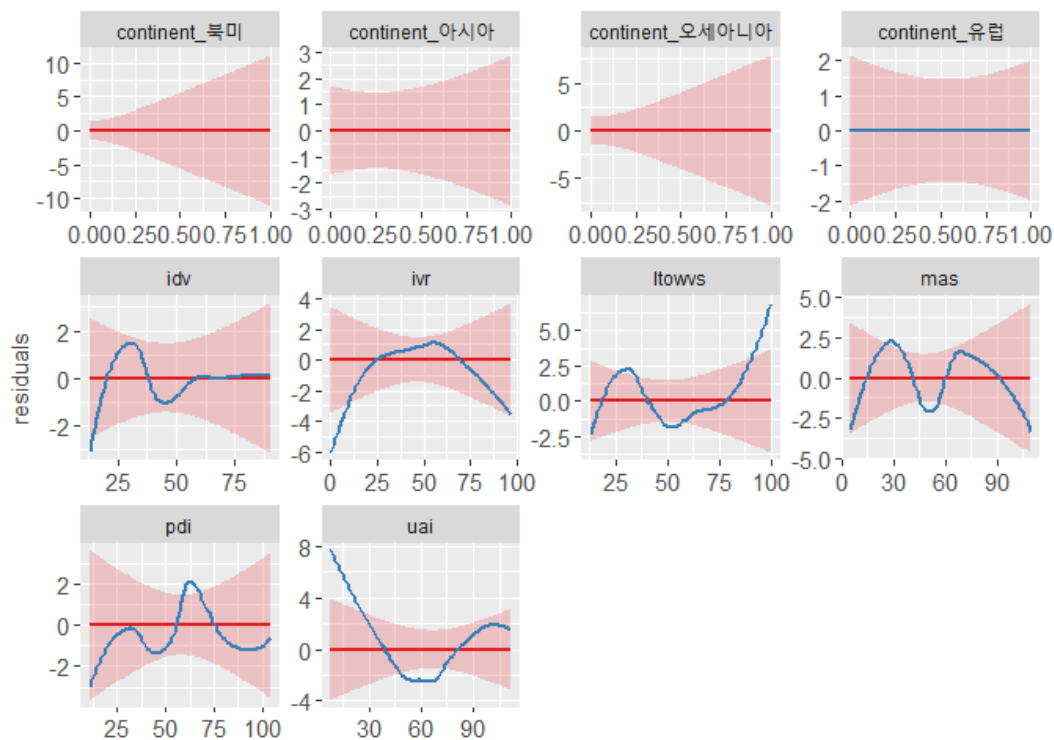
```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```



독립변수별 잔차분포 시각화

```
#>>> resid시각화를 통한 독립변수별 잔차분포 확인
plot_model(reg1,y=,type='resid')
```

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```



다중공선성 진단 및 회귀분석의 가정 검정

```
#>> 'sjPlot' 패키지를 이용한 다중공선성 진단 및 회귀분석 가정 검정
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
```

```
dwtest(reg1) #결과해석 : 회귀분석 가정 검정의 유의성 진단 결과 p-value가 유의수준 0.05보다 작음!
```

```
##
## Durbin-Watson test
##
## data: reg1
## DW = 1.5945, p-value = 0.04401
## alternative hypothesis: true autocorrelation is greater than 0
```


2.5 모델 선택

전진선택법

```
#>모델 선택(예측중심모델) == 1.결측값 사례 제외/ 개선모형과 기존모형결과 객체에 할당 / step 함수를 사용

#>>1. 전진선택법
#(1).결측값 사례 제외 :
complete.cases(main_df) #앞서 결측값을 모두 제거했기에 결측 검사결과 모두 참으로 반환

# (2) 전체 모형과 기본 모형
reg2 = lm(score ~ pdi + idv + mas + uai + ltowvs + ivr + continent_북미 + continent_아시아 +
null = lm(score ~ 1, data=main_df) #기본모형
```

```
# (3) 전진 선택 모형=====
step1 = step(null, scope=list(lower=null, upper=reg2),direction="forward")
```

```
## Error in terms(object): 객체 'null'를 찾을 수 없습니다
```

```
#<<< 1.첫번째 단계의 모형 >>>>
#결과해석 : 첫 번째 단계에서는 기본 모형이 적용되었고, 기본 모형에서 각 독립 변수가 추가되었을 경우
# - AIC 값을 기준으로 살펴보면, 독립변수를 전혀 가정하지 않은 기본 모형에 단체귀속정도(idv)가 추가
# - 단체귀속정도(idv) 변수가 추가되었을 경우에는 기본 모형의 AIC 값보다 낮고, 다른 변수가 추가되었

#<<< 2.두번째 단계의 모형 >>>>
#결과해석 : 두번째 단계에서의 모형은 이전단계에서 선택된 독립변수인 단체귀속정도(idv)가 추가된 모형
# - 독립변수가 추가되었을 경우에 기본모형의 AIC 값보다 낮은 AIC값을 갖는 독립변수는 장기지향성(Lto
# - 이들 변수 중 AIC값이 가장 낮은 독립변수는 장기지향성(Ltowvs)이므로 두번째 단계에서 선택된 독립

#<<< 3.세번째 단계의 모형 >>>
#앞선 단계들과 마찬가지로의 방법으로 살펴본 결과
#과락지향성(ivr)이 세번째 단계에서 선택됨

#<<< 4. 4,5,6 단계의 모형들 >>>
#이후 세개의 단계들도 마찬가지로
#불확실성 회피지수(uai),continent_아시아, 권력거리지수(pdi)가,AIC가 해당 직전단계의 모형보다 작으면
#차례대로 4,5,6단계의 모형을 선택한다.
#이후 AIC값이 낮아지는 독립변수가 없으므로 전진선택방법은 중단된다.
```

#전진 선택 방법을 통한 최종 모형은 독립변수로 자기신뢰감, 부모에대한 애착, 그리고 부정적 양육인

`summary(step1)` #✓ 전진선택법으로 개선한 최종 모형의 결과는 `summary` 함수에 `step1`를 넣어서 확인

Error in summary(step1): 객체 'step1'를 찾을 수 없습니다

#전진선택모델 결과

#- 독립변수인 자기신뢰감, 부모에 대한 애착, 그리고 부정적 양육은 종속변수인 자아존중감에 모두 통계적

최종 모형의 F 값은 44.94 이고, 유의도는 $2.2e-16$ 으로 0.05 보다 낮은수준이므로 최종 모형은 의미가 있

#✓ R^2 값은 0.1868 로 종속변수인 자아존중감의 전체 분산 중에서 3개의 독립변수가 18.68% 를 설명하고 있

3. 결론

해당 보고서에선 세계 혁신지수 를 종속변수로, 홉스테드 6차원문화점수 와 대륙 의 독립변수들이 유의한 영향을 미칠것인지에 대한 가설검정을 연구과제로 일련의 분석과정을 실시해보았고 이 과정을 통해 최종적으로 개선된 모형의 회귀결과를 얻을 수 있었다.

선택한 개선 모형인 전진선택법으로 도출된 다중회귀 결과표에서 알 수 있는 사실은 이렇하다.

개선모델 결과해석

회귀식 :

$$GII\ Score = (0.12 \times idv) + (0.24 \times ltowvs) + (0.23 \times iv) - (0.1 \times uai) - (0.09 \times pdi)$$

- idv , $ltowvs$, ivr 의 세 독립변수는 종속변수에 양의 선형 영향을 미친다.
- uai , pdi 두 독립변수는 종속변수에 음의 선형 영향을 미친다.

- 결과표상 모든 독립변수의 p-value는 유의수준인 0.05보다 작으므로 유의한 영향을 미친다고 할 수 있다.

- 결정계수는 0.74이며 수정결정계수는 0.71이다.
- 더미변수를 활용한 대륙변수의 경우 유럽이 아시아보다 더 큰 영향을 가진다고 할 수 있다.