

2020학년도 2학기  
회귀진단  
기말 레포트

# 세계 행복 지수 분석

# 목차

I. 서론 .....	1
1. 분석 목적 .....	1
II. 본론 .....	1
1. 자료설명 .....	1
1-1. 데이터 설명 .....	1
1-2. 변수 설명 .....	1
1-3. 결측값 처리 .....	2
1-4. 요약 통계량 .....	2
1-5. 변수그림 .....	3
1-6. 상관 분석 .....	15
2. 데이터 분석 .....	16
2-1. 회귀분석 .....	16
2-2. 연도별 비교 .....	24
2-3. 한국의 행복점수 .....	27
III. 결론 .....	28
1. 기대효과 .....	28
2. 한계점 .....	28
R코드 .....	29
참고문헌 .....	38

# I. 서론

## 1. 분석 목적

세계 행복 보고서는 유엔 산하 자문기구인 SDSN이 국제 행복의 날인 매년 3월 20일에 발표하는 보고서이다. SDSN은 1인당 국내총생산(GDP), 사회적 지원, 기대 수명, 사회적 자유, 관용, 부정 부패, 미래에 대한 불안감 이렇게 총 7가지 지표를 기준으로 국가별 행복지수를 산출했다.

세계 행복 보고서를 통해 행복 지수가 높은 국가들이 행복 지수의 지표인 GDP, 사회적 지원 등 여러 기준에서 모두 높은 지, 행복 지수와 기준 변수들 간의 관계를 알아보고자 한다. 더 나아가 행복 지수의 상위 10개국뿐만 아니라 OECD 회원국 중 압도적으로 자살률 1위인 한국은 여러 기준에서 하위권에 차지하고 있는지 등을 알아보고자 한다.

# II. 본론

## 1. 자료 설명

### 1-1. 데이터 설명

- 머신 러닝과 데이터 사이언스 커뮤니티인 Kaggle 사이트에서 세계 행복 보고서에 관한 데이터를 사용하였다. 2015년부터 2020년까지 자료가 있지만, 이 중 2017, 2018, 2019년의 데이터만 사용하기로 하였다. 2020년의 최신 정보는 변수의 종류가 전년도와 다른 부분이 많아, 데이터 개수를 늘리기 위해 변수 종류가 비슷한 3개년의 데이터를 사용하였다.

- 2017년 데이터 개수는 155개, 2018년 데이터 개수는 156개, 2019년 데이터 개수는 156개로 최종 데이터 개수는 467개이다.

### 1-2. 변수 설명

<표 1> 변수 설명

변수 이름	변수 설명	변수 종류	세부사항
rank	행복 순위	연속형	행복 점수가 가장 높은 국가나 지역부터 측정한 순위
region	국가, 지역	문자형	156개의 국가나 지역
gdp	GDP	연속형	1인당 국내총생산(GDP)가 행복 지수 계산에 기여한 정도
support	사회적 지원	연속형	사회적 지원이 행복 지수 계산에 기여한 정도

life	기대 수명	연속형	기대 수명이 행복 지수 계산에 기여한 정도
freedom	자유	연속형	자유가 행복 지수 계산에 기여한 정도
generosity	관용	연속형	관용이 행복 지수 계산에 기여한 정도
corruption	부정부패 인식	연속형	부정부패 인식이 행복 지수 계산에 기여한 정도
year	연도	범주형	2017년 = 2 2018년 = 1 2019년 = 0
y	행복 점수	연속형	10이 가장 행복한 점수라고 할 경우 0에서 10까지의 점수로, 자신의 행복을 어떻게 평가하시겠습니까? 라는 질문을 통해 측정한 항목

### 1-3. 결측값 처리

- 2018년 데이터에서 generosity 결측치를 generosity의 평균인 0.112로 대체하였다.

### 1-4. 요약 통계량

#### 1-4-1. 2017년

<표 2> 2017년 자료 요약 통계량

	y	gdp	support	life	freedom	generosity	corruption
최솟값	2.6930	0.0000	0.00000	0.0000000	0.0000000	0.0000000	0.0000000
제 1사분위수	4.5055	0.66337	1.042635	0.3698663	0.3036772	0.1541061	0.0572707
중앙값	5.2790	1.06458	1.253918	0.6060416	0.4374543	0.2315385	0.0898475
평균	5.35402	0.98472	1.188898	0.5513408	0.408786	0.2468835	0.1231202
제 3사분위수	6.1015	1.31803	1.414316	0.7230075	0.5165614	0.3237623	0.1532963
최댓값	7.5370	1.87077	1.610574	0.9494924	0.6582487	0.8380752	0.4643078

#### 1-4-2. 2018년

<표 3> 2018년 자료 요약 통계량

	y	gdp	support	life	freedom	generosity	corruption
최솟값	2.90500	0.00000	0.00000	0.00000	0.0000	0.000	0.0000
제 1사분위수	4.45375	0.61625	1.06675	0.42225	0.3560	0.1095	0.0510
중앙값	5.37800	0.94950	1.25500	0.64400	0.4870	0.1740	0.0820

평균	5.37592	0.89145	1.21324	0.59735	0.4545	0.18101	0.112
제 3사분위수	6,16750	1.19775	1.46300	0.77725	0.5785	0.2390	0.1365
최댓값	7.63200	2.09600	1.64400	1.03000	0.7240	0.5980	0.4570

#### 1-4-3. 2019년

<표 4> 2019년 자료 요약 통계량

	y	gdp	support	life	freedom	generosity	corruption
최솟값	2.8530	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
제 1사분위수	4.5445	0.60275	1.05575	0.54775	0.30800	0.10875	0.04700
중앙값	5.3795	0.96000	1.27150	0.78900	0.41700	0.17750	0.08550
평균	5.4071	0.90515	1.20881	0.72524	0.39257	0.18485	0.11060
제 3사분위수	6.1845	1.23250	1.45250	0.88175	0.50725	0.24825	0.14125
최댓값	7.7690	1.68400	1.62400	1.14100	0.63100	0.56600	0.45300

#### 1-4-4. 2017, 2018, 2019년 합친 전체 데이터

<표 5> 2017, 2018, 2019년 합친 자료 요약 통계량

	y	gdp	support	life	freedom	generosity	corruption
최솟값	2.6930	0.0000000	0.000000	0.0000000	0.0000000	0.0000000	0.0000000
제 1사분위수	4.5115	0.6282034	1.051500	0.4547138	0.3130000	0.1200000	0.0520000
중앙값	5.3390	0.9860000	1.258190	0.6570000	0.4433235	0.1901336	0.0867232
평균	5.379064	0.9269814	1.203681	0.6248005	0.418642	0.204154	0.1152241
제 3사분위수	6.1705	1.2450000	1.440726	0.8175000	0.5355000	0.2625000	0.1435000
최댓값	7.7690	2.0960000	1.644000	1.1410000	0.7240000	0.8380752	0.4643078

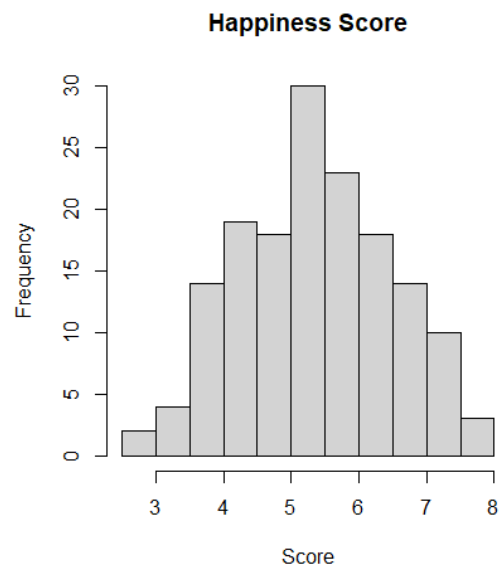
### 1.5 변수 그림

#### 1-5-1. 2017년

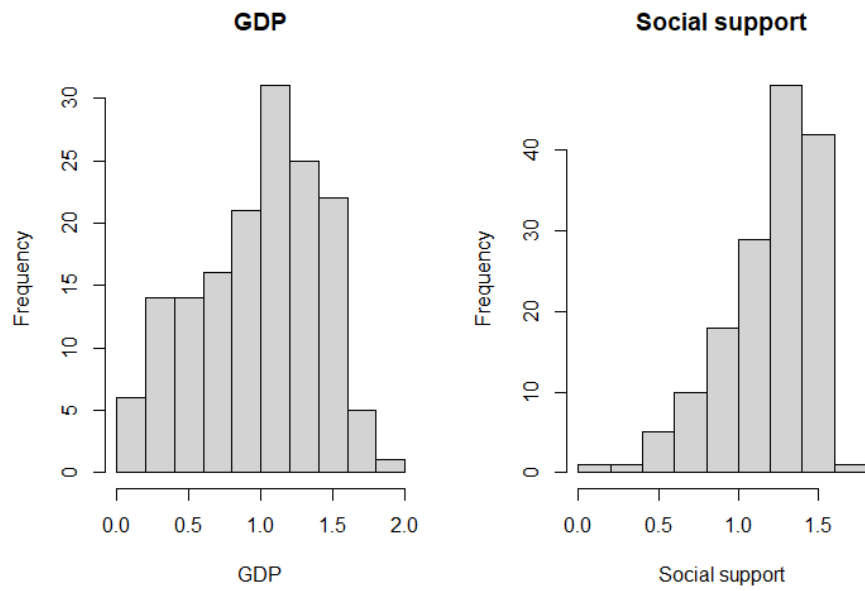
##### 1) 변수들에 대한 히스토그램

- y(score), gdp, social support, life, freedom 변수의 히스토그램은 중심이 솟은 종 모양 그래프임을 알 수 있다. 이에 반해 generosity, corruption 변수의 히스토그램은 왼쪽으로 치우친 그래프임을 알 수 있다.

<그림 1> 2017년 행복 점수(y) 히스토그램



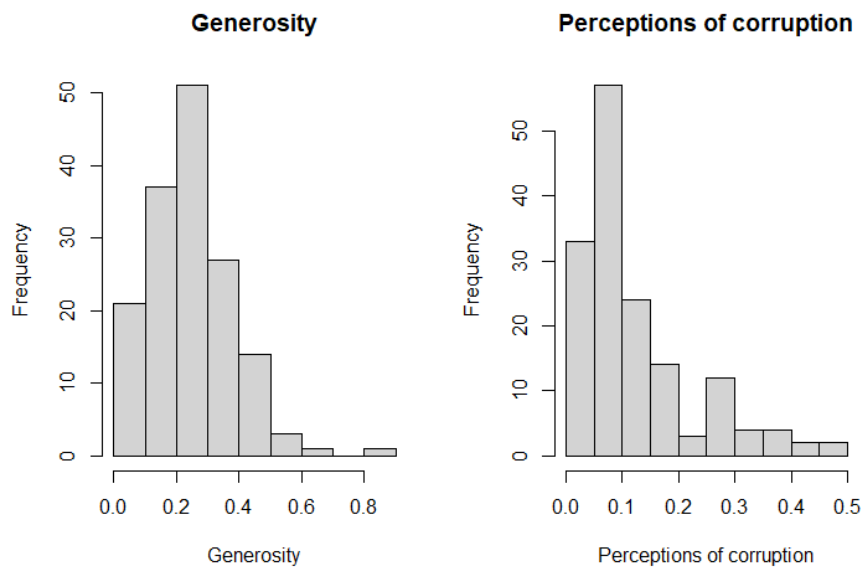
<그림 2> 2017년 gdp, social support 히스토그램



<그림 3> 2017년 life, freedom 히스토그램



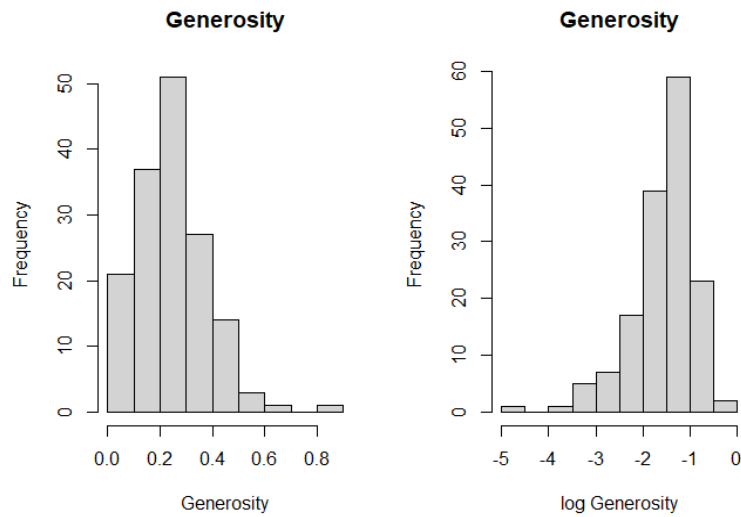
<그림 4> 2017년 generosity, corruption 히스토그램



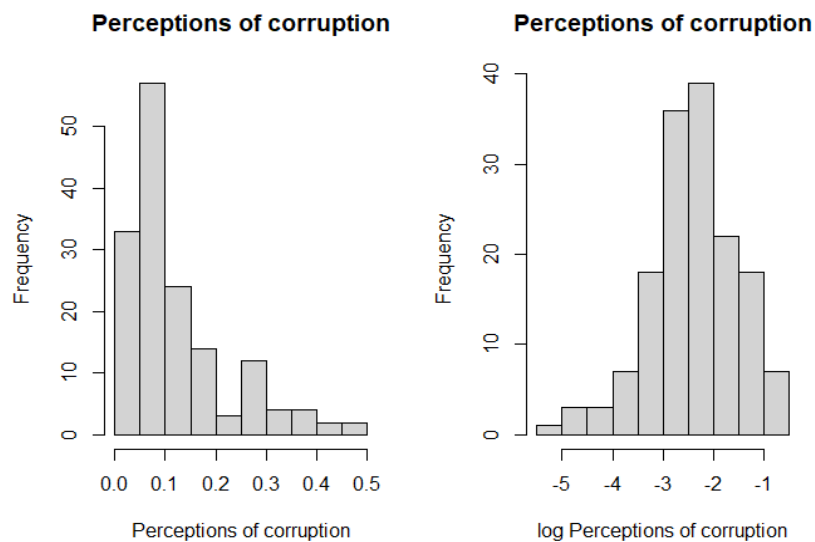
## 2) 로그 변환

- 변수 generosity, corruption를 로그 변환했을 경우 왼쪽으로 치우치던 그래프가 가운데가 솟은 그래프로 바뀌기 때문에 로그 변환을 하였다.

<그림 5> 2017년 로그 변환한 generosity 히스토그램



<그림 6> 2017년 로그 변환한 corruption 히스토그램



## 1-5-2. 2018년

### 1) 변수들에 대한 히스토그램

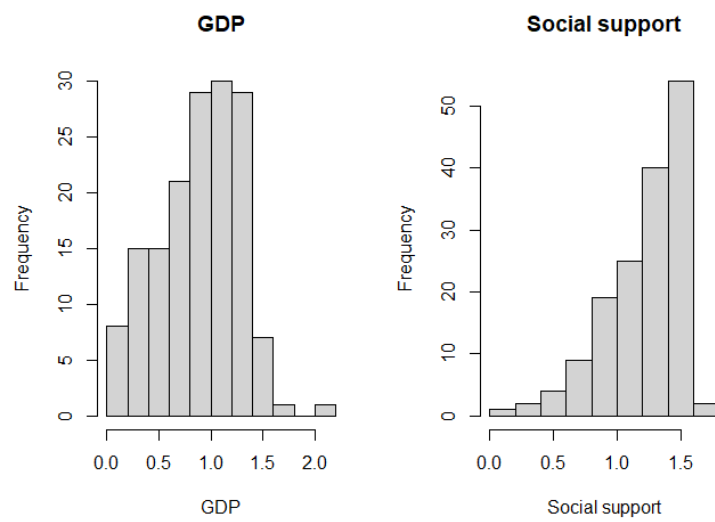
- 그림 7, 그림 8, 그림 9, 그림 10을 보면 y(score), gdp, social support, life, freedom 변수의 히스토그램이 가운데가 솟은 그래프이다. 이에 반해 generosity, corruption 변수의 히스토그램은 왼쪽으로 치우친 그래프이다.



<그림 7> 2018년 행복 점수(y) 히스토그램



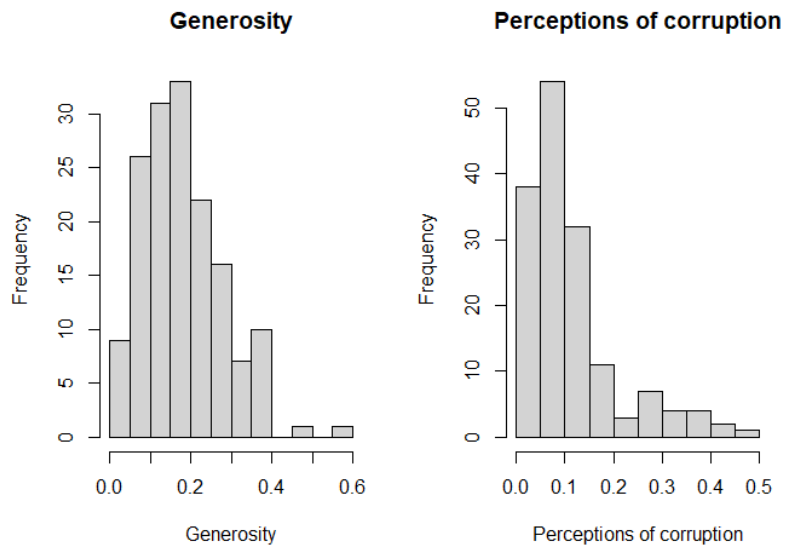
<그림 8> 2018년 gdp, social support 히스토그램



<그림 9> 2018년 life, freedom 히스토그램



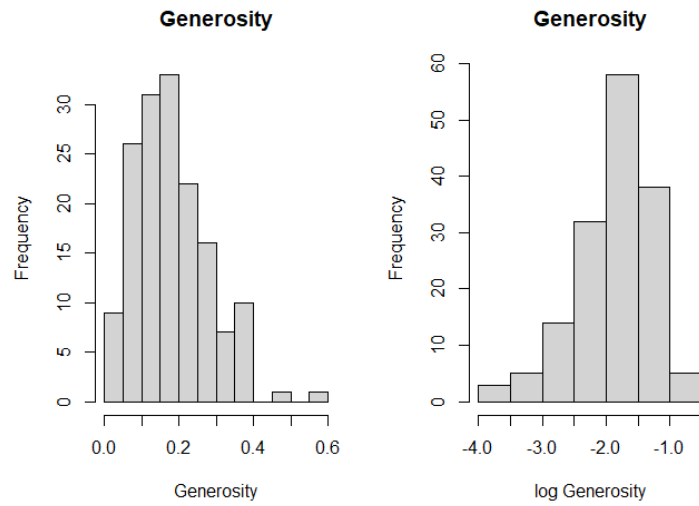
<그림 10> 2018년 generosity, corruption 히스토그램



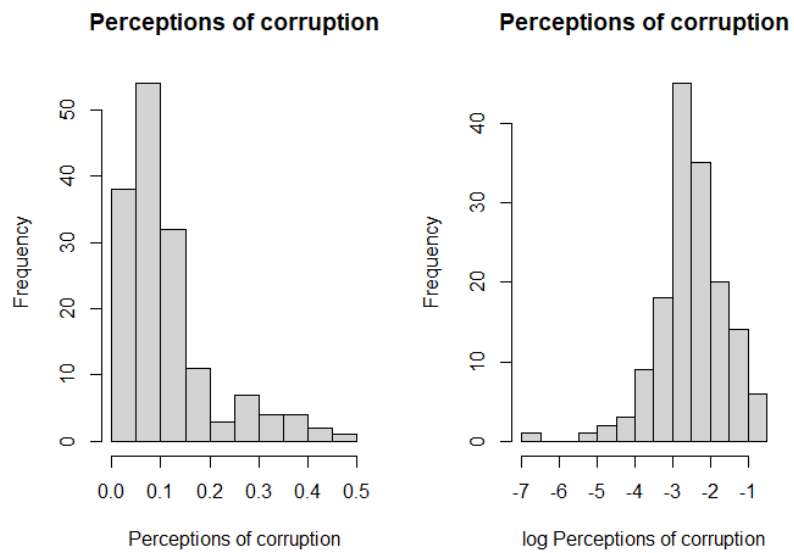
## 2) 로그변환

- 변수 generosity, corruption이 왼쪽으로 치우쳐 있어 로그 변환을 통해 가운데가 솟은 종 모양 그래프가 된 것으로 바뀌었다.

<그림 11> 2018년 로그 변환한 generosity 히스토그램



<그림 12> 2018년 로그 변환한 corruption 히스토그램



### 1-5-3. 2019년

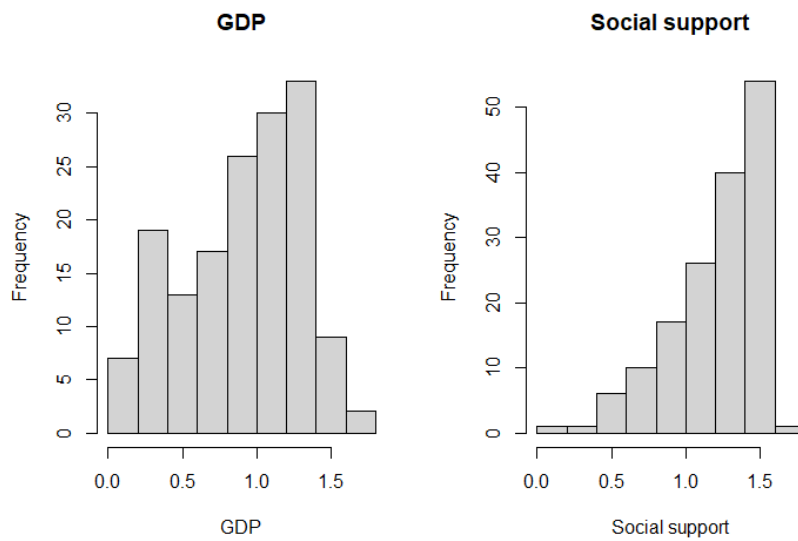
#### 1) 변수들에 대한 히스토그램

- y(score), gdp, social support, life, freedom 변수의 히스토그램을 보면 중심이 솟은 종 모양 그래프이지만, generosity, corruption 변수의 히스토그램은 한쪽으로 치우친 그래프로 나타났다.

<그림 13> 2019년 행복 점수 히스토그램



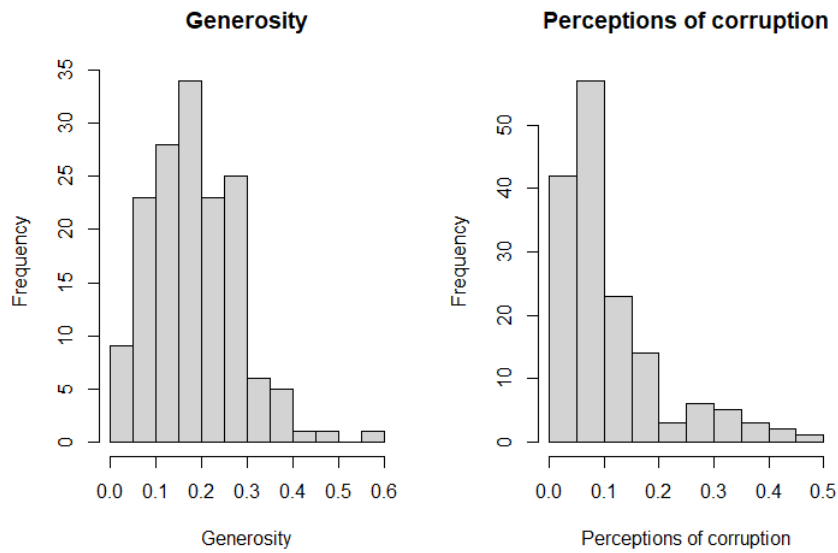
<그림 14> 2019년 gdp, social support 히스토그램



<그림 15> 2019년 life, freedom 히스토그램



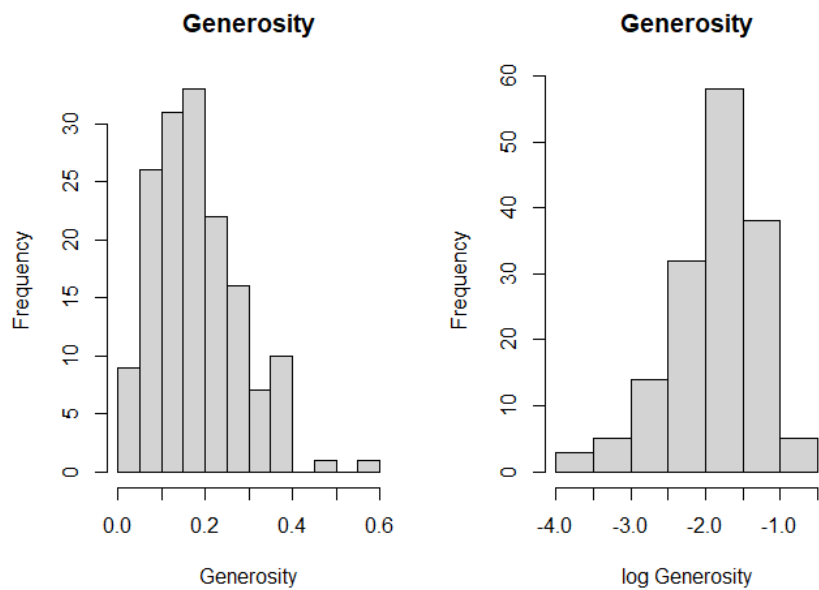
<그림 16> 2019년 generosity, corruption 히스토그램



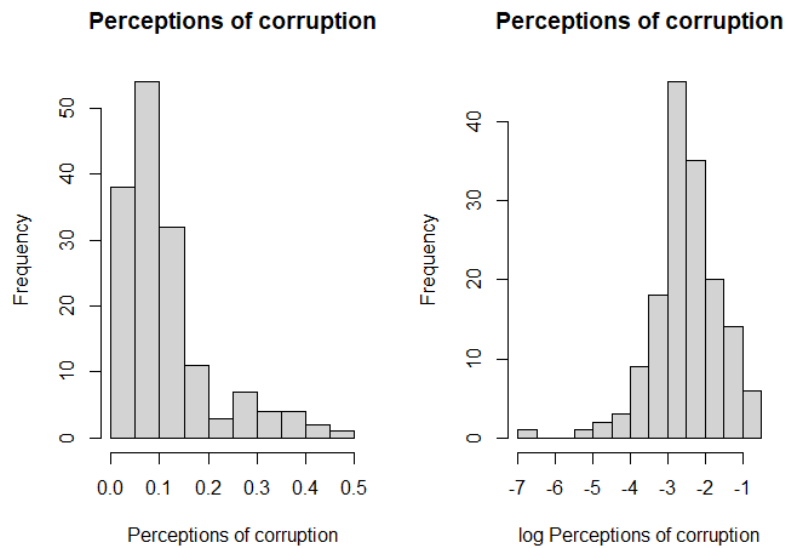
## 2) 로그 변환

- 변수 generosity, corruption를 로그 변환한다면 왼쪽으로 치우치던 그래프가 중심이 솟은 그래프로 바뀌기 때문에 로그 변환을 해야 한다.

<그림 17> 2019년 로그 변환한 generosity 히스토그램



<그림 18> 2019년 로그 변환한 corruption 히스토그램



#### 1-5-4. 2017, 2018, 2019년 합친 전체 데이터

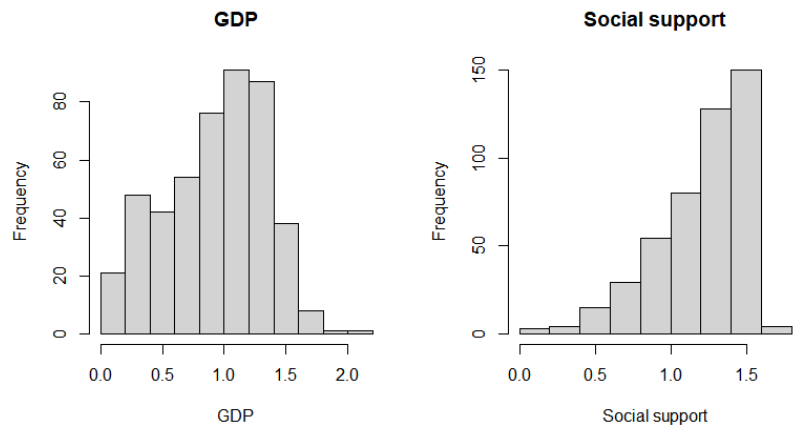
##### 1) 변수들에 대한 히스토그램

- y(score), gdp, social support, life, freedom 변수의 히스토그램은 중심이 솟은 종 모양 그래프임을 알 수 있다. 이에 반해 그림 22를 보면 generosity, corruption 변수의 히스토그램은 왼쪽으로 치우친 그래프이기 때문에 변환해야 할 필요가 있다.

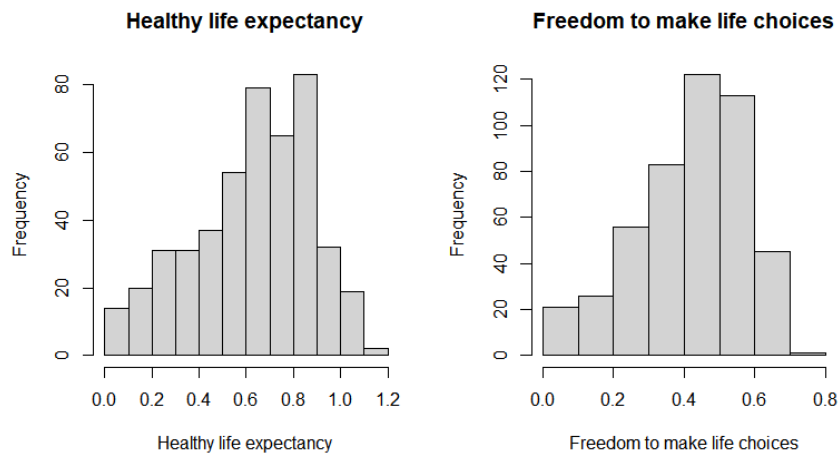
<그림 19> 2017, 2018, 2019년 합친 데이터의 행복 점수 히스토그램



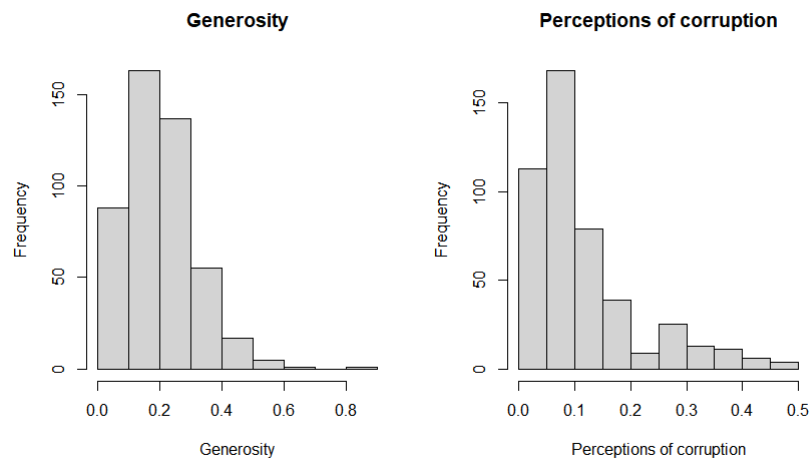
<그림 20> 2017, 2018, 2019년 합친 데이터의 gdp, social support 히스토그램



<그림 21> 2017, 2018, 2019년 합친 데이터의 life, freedom 히스토그램



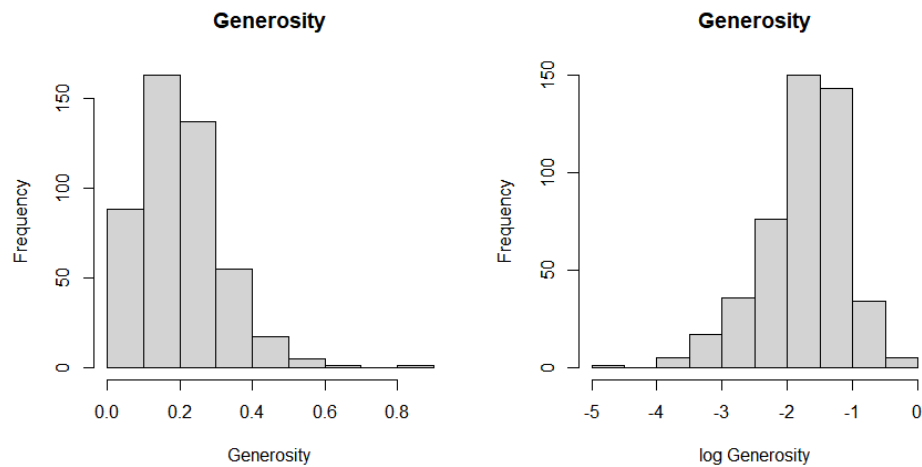
<그림 22> 2017, 2018, 2019년 합친 데이터의 generosity, corruption 히스토그램



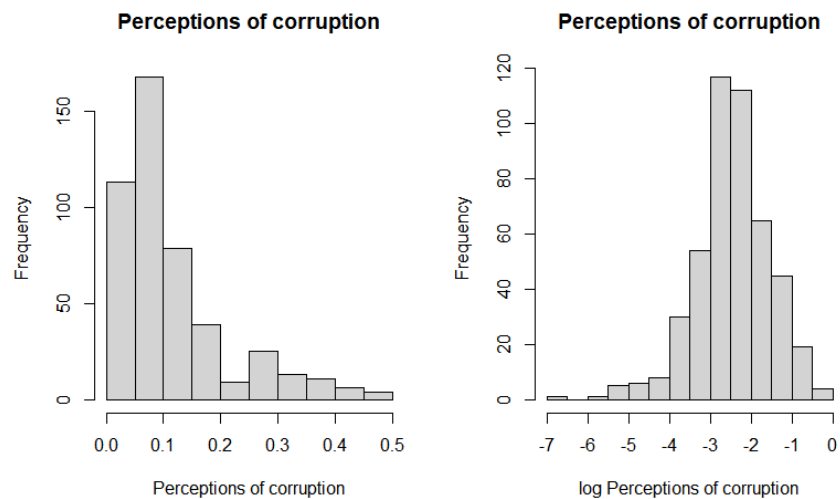
## 2) 로그 변환

- 변수 generosity, corruption를 로그 변환했을 경우 왼쪽으로 치우치던 그래프가 가운데가 솟은 그래프로 바뀌기 때문에 로그 변환을 한 데이터를 사용하였다.

<그림 23> 2017, 2018, 2019년 합친 데이터의 로그 변환한 generosity 히스토그램



<그림 24> 2017, 2018, 2019년 합친 데이터의 로그 변환한 corruption 히스토그램

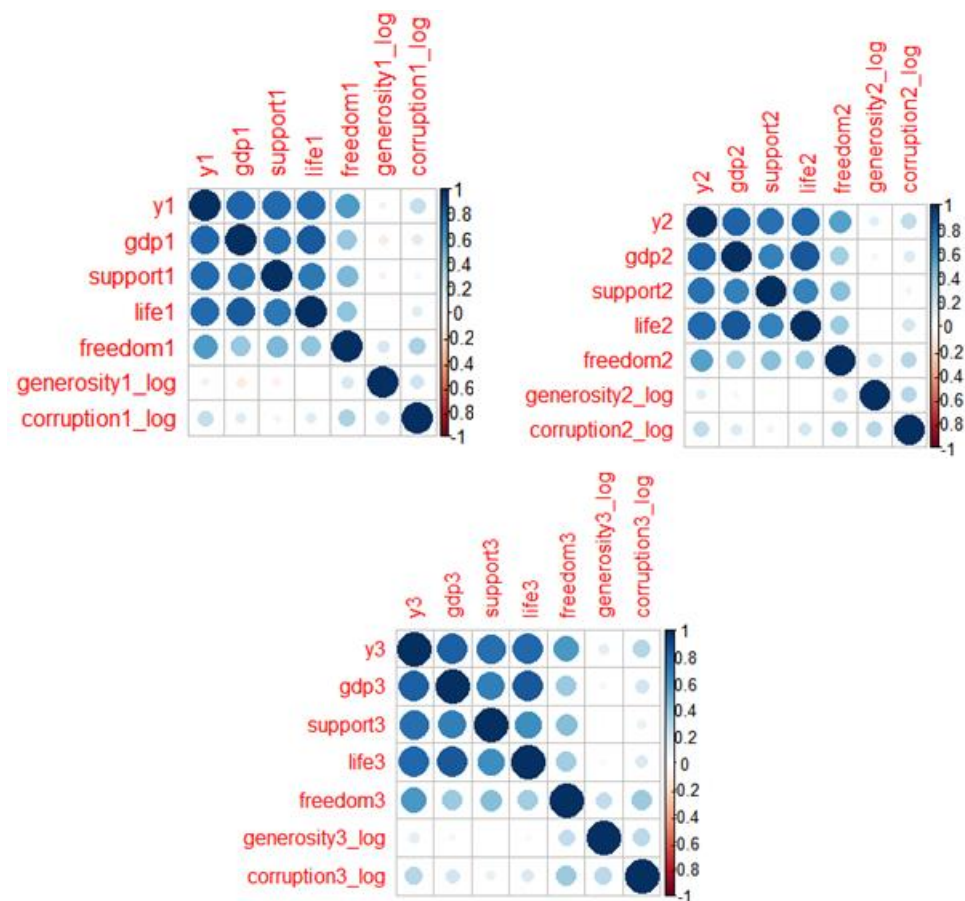




## 1-6. 상관분석

- 상관분석을 통해 행복 점수(y)와 다른 독립 변수간 관계성이 있는지 분석하였다. 연도별로 행복 점수(y)와 각 변수들 간의 관계를 알아보았고, y1은 2017년 행복 점수, y2는 2018년 행복 점수, y3은 2019년 행복 점수를 의미한다. 분석 결과 모두 상관계수가 양수로 정비례 관계라는 것을 파악하였다. 하지만 행복 점수(y)와 generosity 간에 양의 상관관계 다른 변수들에 비해 약하다는 것을 알 수 있다.

<그림 25> 상관행렬 히트맵



- 그림 25는 상관행렬 히트맵으로, 상관계수가 클수록 원의 크기가 크고 색깔이 진하게 나오며, 상관계수가 양수이면 파란색 음수면 빨간색 계열로 표시가 된다.

## 2. 데이터 분석

### 2-1. 회귀분석

#### 1) 모형 선택

회귀분석 중에서 다중선형회귀분석을 이용하였고 가변수의 유무와 변수변환의 유무에 따라 총 4개의 모형을 적합하였고 모형은 다음과 같다.

$$Model\ 1 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 D_1 + \beta_8 D_2 + \epsilon$$

$$Model\ 2 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 \log(X_6) + \beta_7 D_1 + \beta_8 D_2 + \epsilon$$

$$Model\ 3 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 \log(X_5) + \beta_6 X_6 + \beta_7 D_1 + \beta_8 D_2 + \epsilon$$

$$Model\ 4 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 \log(X_5) + \beta_6 \log(X_6) + \beta_7 D_1 + \beta_8 D_2 + \epsilon$$

<표 6> 회귀분석 모형에서 변수 설명

변수	변수 설명
$X_1$	gdp
$X_2$	support
$X_3$	life
$X_4$	freedom
$X_5$	generosity
$X_6$	corruption
$D_1$	year(2018)
$D_2$	year(2017)

이러한 모형 4개를 이용하여 회귀분석을 실시하여 최적의 모형을 선택하도록 하였다. 연도별 데이터를 모아 추정하였기 때문에, 가변수인 연도 변수(year)와 다른 독립변수들 간에 교호작용이 없는 경우로 가정하였다. 이 때 분산팽창인자(Variance Inflation Factor, VIF)를 구하여 모형별로 다중공선성이 있는지 판단하였다.

<표 7> 모형 1의 회귀분석 결과

모형	변수	추정값 (표준오차)	검정통계량 (유의확률)	VIF	$R^2$	$Adjust\ R^2$
1	절편	1.80776 (0.11859)	15.243 ( $< 2e-16$ )		0.7927	0.7891
	$X_1$	0.89978 (0.11924)	7.546 ( $2.44e-13$ )	2.028702		
	$X_2$	1.06184 (0.12059)	8.805 ( $< 2e-16$ )	1.498782		
	$X_3$	1.05047 (0.18607)	5.646 ( $2.89e-08$ )	1.975262		
	$X_4$	1.42595 (0.19629)	7.265 ( $1.62e-12$ )	1.269966		
	$X_5$	0.48738 (0.23594)	2.066 (0.03942)	1.136716		
	$X_6$	0.81024 (0.29401)	2.756 (0.00609)	1.203347		
	$D_1$	0.02322 (0.06386)	0.364 (0.71626)	1.142223		
	$D_2$	0.01565 (0.07234)	0.216 (0.82878)			

<표 8> 모형 2의 회귀분석 결과

모형	변수	추정값 (표준오차)	검정통계량 (유의확률)	VIF	$R^2$	$Adjust R^2$
2	절편	2.06882 (0.14930)	13.857 ( $< 2e-16$ )		0.7934	0.7898
	$X_1$	0.92811 (0.11750)	7.899 ( $2.10e-14$ )	2.002539		
	$X_2$	1.06549 (0.12033)	8.855 ( $< 2e-16$ )	1.498196		
	$X_3$	1.04357 (0.18578)	5.617 ( $3.37e-08$ )	1.975660		
	$X_4$	1.47282	7.766	1.229150		

		(0.18964)	(5.34e-14)			
	$X_5$	0.51236 (0.23250)	2.204 (0.0280)	1.122088		
	$\log(X_6)$	0.08704 (0.02863)	3.041 (0.0025)	1.089058		
	$D_1$	0.01728 (0.06368)	0.271 (0.7862)	1.141616		
	$D_2$	0.00914 (0.07217)	0.127 (0.8993)			

<표 9> 모형 3의 회귀분석 결과

모형	변수	추정값 (표준오차)	검정통계량 (유의확률)	VIF	$R^2$	$Adjust\ R^2$
3	절편	2.132681 (0.140738)	15.154 ( $< 2e-16$ )		0.7953	0.7917
	$X_1$	0.930287 (0.119023)	7.816 (3.77e-14)	2.037933		
	$X_2$	1.060346 (0.119816)	8.850 ( $< 2e-16$ )	1.498672		
	$X_3$	1.018816 (0.185346)	5.497 (6.43e-08)	1.980182		
	$X_4$	1.433344 (0.192319)	7.453 (4.59e-13)	1.252254		
	$\log(X_5)$	0.127948 (0.040064)	3.194 (0.0015)	1.094495		
	$X_6$	0.736103 (0.291583)	2.525 (0.0119)	1.201052		
	$D_1$	0.022228 (0.063419)	0.350 (0.7261)	1.133501		
	$D_2$	0.005485 (0.071357)	0.077 (0.9388)			

<표 10> 모형 4의 회귀분석 결과

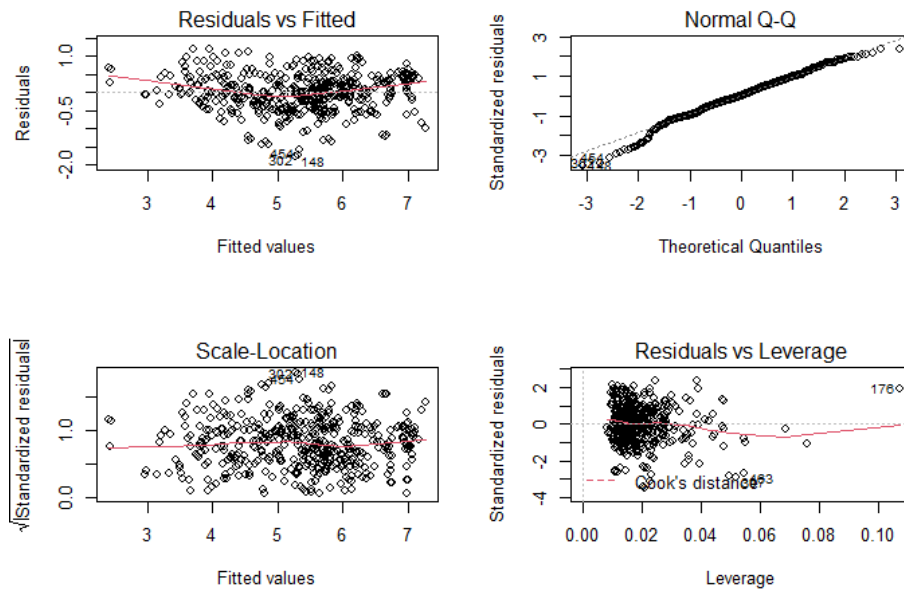
모형	변수	추정값 (표준오차)	검정통계량 (유의확률)	VIF	$R^2$	$Adjust\ R^2$
4	절편	2.3809815 (0.1548533)	15.376 ( $< 2e-16$ )		0.796	0.7925
	$X_1$	0.9551699 (0.1172427)	8.147 (3.58e-15)	2.010922		
	$X_2$	1.0641741 (0.1195503)	8.901 ( $< 2e-16$ )	1.497936		
	$X_3$	1.0125041 (0.1850497)	5.472 (7.36e-08)	1.980433		
	$X_4$	1.4768596 (0.1847582)	7.993 (1.08e-14)	1.205101		
	$\log(X_5)$	0.1299224 (0.0396141)	3.280 (0.00112)	1.084064		
	$\log(X_6)$	0.0804768 (0.0284886)	2.825 (0.00494)	1.090657		
	$D_1$	0.0166805 (0.0632383)	0.264 (0.79207)	1.132889		
	$D_2$	0.0003414 (0.0712068)	0.005 (0.99618)			

- 모형 1, 모형 2, 모형 3, 모형 4에서 모든 독립변수(가변수 포함)들의 VIF가 10이하이기 때문에 다중공선성의 문제가 없다고 판단된다.

## 2) 모형 진단

(1)  $Model\ 1 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 D_1 + \beta_8 D_2 +$

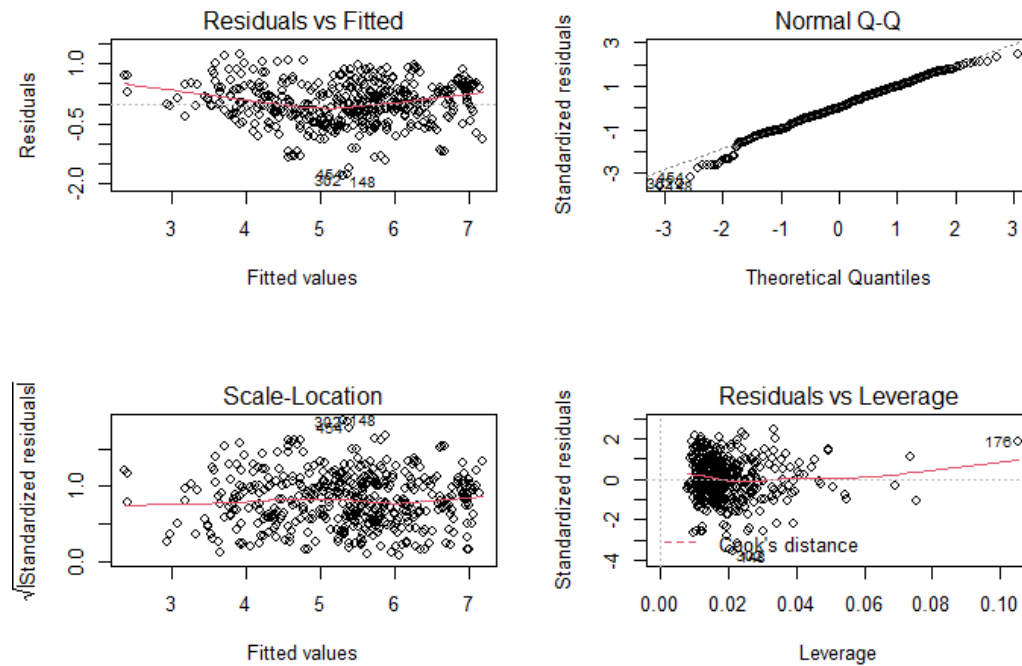
<그림 26> 모형 1의 모형진단



- 그림 26에서 왼쪽 위 그래프에서 잔차와 예측값 사이에 어떠한 관계가 있다고 판단되지 않기 때문에 선형성 가정을 만족한다고 할 수 있다.
- 그림 26에서 Normal Q-Q plot을 보면 그래프의 점들이 직선에 모여 있기 때문에 잔차의 정규성 가정을 만족한다고 할 수 있다.
- 그림 26에서 왼쪽 아래 그래프를 보면 수평선 주위의 일정한 범위 내에서 랜덤하게 나타나기 때문에 등분산성 가정을 만족한다고 할 수 있다.
- 그림 26에서 오른쪽 아래 그래프를 통해 숫자가 써 있는 점이 이상치라는 것을 알 수 있다.
- 따라서 모형 1은 선형성 가정, 잔차의 정규성 가정, 등분산성 가정을 만족한다고 할 수 있다.

(2) Model 2 :  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 \log(X_6) + \beta_7 D_1 + \beta_8 D_2 + \epsilon$

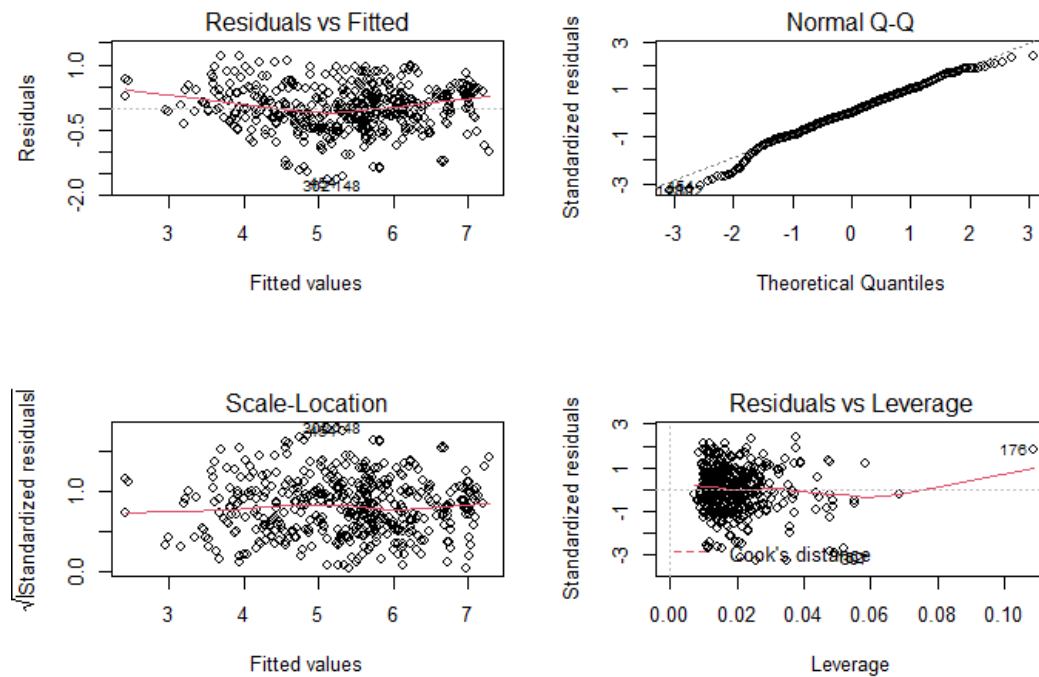
<그림 27> 모형 2의 모형진단



- 그림 27에서 왼쪽 위 그래프에서 점들이 랜덤하게 나타났기 때문에 잔차와 예측값 사이에 어떠한 관계가 있다고 판단되지 않아 선형성 가정을 만족한다고 할 수 있다.
- 그림 27에서 오른쪽 위 그래프인 Normal Q-Q plot을 보면 그래프의 점들이 직선에 모여 있어 잔차의 정규성 가정을 만족한다고 할 수 있다.
- 그림 27에서 왼쪽 아래 그래프를 보면 수평선 주위의 일정한 범위 내에서 랜덤하게 나타나기 때문에 등분산성 가정을 만족한다고 할 수 있다.
- 그림 27에서 오른쪽 아래 그래프를 보면 숫자가 써 있는 점이 이상치라는 것을 알 수 있다.
- 따라서 모형 2는 선형성 가정, 잔차의 정규성 가정, 등분산성 가정을 만족한다.

(3) Model 3 :  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 \log(X_5) + \beta_6 X_6 + \beta_7 D_1 + \beta_8 D_2 + \epsilon$

<그림 28> 모형 3의 모형진단

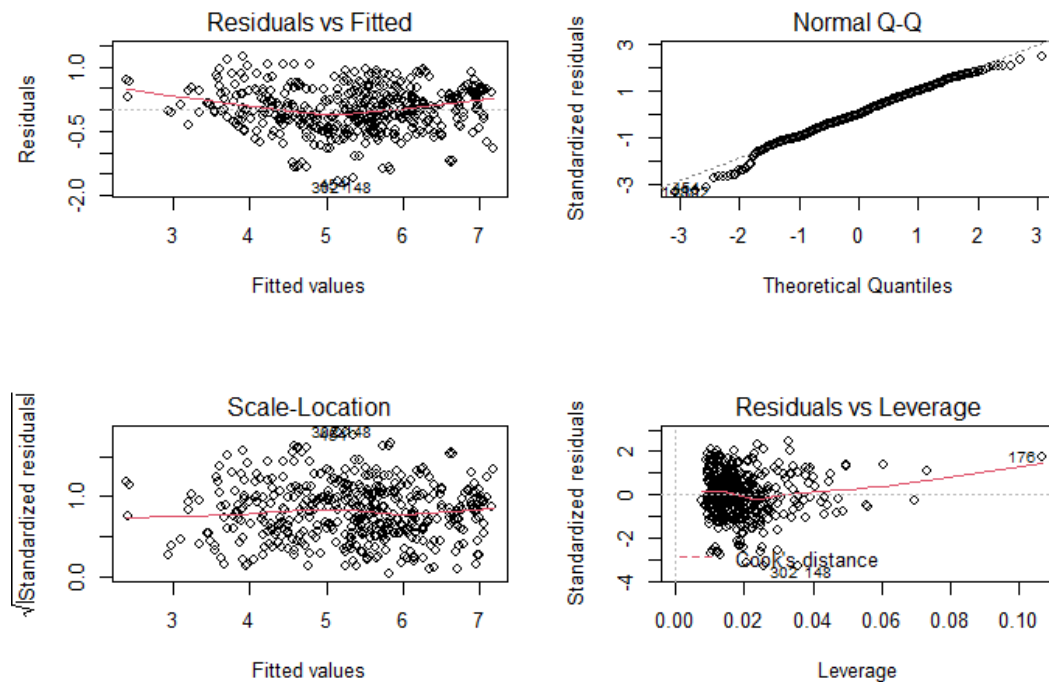


- 그림 28에서 왼쪽 위 그래프에서 점들이 랜덤하게 나타났기 때문에 선형성 가정을 만족한다고 판단된다.
- 그림 28에서 오른쪽 위 그래프인 Normal Q-Q plot을 보면 그래프의 점들이 직선에 모여 있어 잔차의 정규성 가정을 만족한다고 할 수 있다.
- 그림 28에서 왼쪽 아래 그래프를 보면 수평선 주위의 일정한 범위 내에서 랜덤하게 나타나 등분산성 가정을 만족한다고 할 수 있다.
- 그림 28에서 오른쪽 아래 그래프를 보면 숫자가 써 있는 점이 이상치라는 것을 알 수 있다.
- 따라서 모형 3은 선형성 가정, 잔차의 정규성 가정, 등분산성 가정을 만족한다.



(4) Model 4 :  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 \log(X_5) + \beta_6 \log(X_6) + \beta_7 D_1 + \beta_8 D_2 + \epsilon$

<그림 29> 모형 4의 모형진단



- 그림 29에서 왼쪽 위 그래프에서 점들이 랜덤하게 나타났기 때문에 선형성 가정을 만족한다고 판단된다.
- 그림 29에서 오른쪽 위 그래프인 Normal Q-Q plot을 보면 그래프의 점들이 직선에 모여 있어 잔차의 정규성 가정을 따른다고 할 수 있다.
- 그림 29에서 왼쪽 아래 그래프를 보면 수평선 주위의 일정한 범위 내에서 랜덤하게 나타나 등분산성 가정을 만족한다고 판단된다.
- 그림 29에서 오른쪽 아래 그래프를 보면 숫자가 써 있는 점이 이상치라는 것을 알 수 있다.
- 따라서 모형 4는 선형성 가정, 잔차의 정규성 가정, 등분산성 가정을 만족한다.

### 3) 모형 선택 결과

- 모형진단의 결과와 표 7, 표 8, 표 9, 표 10에서 수정결정계수를 토대로 수정결정계수가 가장 높은 모형인 모형 4를 선택하였다.

<표 11> 모형별 수정결정계수

모형	모형 1	모형 2	모형 3	모형 4
수정결정계수	0.7891	0.7898	0.7917	0.7925

- 지렛값을 이용하여 독립변수 기준으로 특이값은 없다는 것을 알 수 있다. 종속변수 기준으로 특이값은 표준화제외잔차를 이용하여 판별하였고 다음 표 11에 정리하였다.

<표 12> 최종 모형인 모형 4에서의 종속변수 기준으로 특이값

번호	67	102	231	254	368	392	405
y(score)	2.102027	2.338124	2.062976	2.007978	2.092676	2.066117	2.482676

- 영향력이 큰 관측값은 cook's distance를 이용하여 판별하였을 때 없다는 결과를 얻었다.
- 또한 단계별 회귀방법을 모형 4에 적용하였을 때, 모형 4 그대로 사용하는 것이 좋다는 결과를 얻었다.
- 회귀계수를 본 결과 행복 점수(y)에 영향을 끼치는 정도는 사회적 지원(support), 1인당 국내총생산(gdp), 자유(freedom), 기대 수명(life), 관용(generosity), 부정부패인식(corruption) 순으로 나타난 것을 알 수 있다. 즉, 행복 점수에 가장 영향을 끼치는 변수는 사회적 지원(support)이다. 따라서 사회적 지원이 많을수록 행복 점수가 높아질 수 있다고 할 수 있다.

## 2-2. 연도별 비교

### 2-2-1. 2017년

#### 1) 각 변수 별 상위 10개 국가(지역)

<표 13> 2017년 각 변수 별 상위 10개 국가(지역)

rank	y(score)	gdp	support	life	freedom	generosity	corruption
1	Norway	Qatar	Iceland	Singapore	Uzbekistan	Myanmar	Singapore
2	Denmark	Luxembourg	Ireland	Hong Kong S.A.R., China	Norway	Indonesia	Rwanda
3	Iceland	Singapore	Denmark	Japan	Cambodia	Malta	Qatar
4	Switzerland	Kuwait	Uzbekistan	South Korea	Iceland	Thailand	Denmark

5	Finland	United Arab Emirates	New Zealand	Spain	Denmark	New Zealand	Sweden
6	Netherlands	Norway	Finland	Switzerland	Switzerland	Syria	New Zealand
7	Canada	Switzerland	Norway	Italy	Finland	United Kingdom	Finland
8	New Zealand	Hong Kong S.A.R., China	Spain	Luxembourg	New Zealand	Sri Lanka	Switzerland
9	Sweden	United States	Switzerland	Cyprus	Sweden	Haiti	United Arab Emirates
10	Australia	Ireland	Australia	France	Canada	Australia	Luxembourg

- 행복 점수를 기준으로 Norway, Denmark, Iceland, Switzerland, Finland, Netherlands, Canada, New Zealand, Sweden, Australia가 상위 10개 국가(지역)로 선정되었다.

## 2-2-2. 2018년

### 1) 각 변수 별 상위 10개 국가(지역)

<표 14> 2018년 각 변수 별 상위 10개 국가(지역)

rank	y(score)	gdp	support	life	freedom	generosity	corruption
1	Finland	United Arab Emirates	Iceland	Hong Kong	Uzbekistan	Myanmar	Singapore
2	Norway	Qatar	New Zealand	Singapore	Cambodia	Indonesia	Rwanda
3	Denmark	Luxembourg	Finland	Japan	Norway	Haiti	Denmark
4	Iceland	Singapore	Denmark	Spain	Denmark	Malta	Finland
5	Switzerland	Kuwait	Uzbekistan	South Korea	Finland	Syria	New Zealand
6	Netherlands	Norway	Ireland	Italy	Iceland	New Zealand	Sweden
7	Canada	Ireland	Norway	Switzerland	Somalia	Thailand	Switzerland
8	New Zealand	Switzerland	Australia	Iceland	New Zealand	Bhutan	Norway
9	Sweden	Hong Kong	Israel	Sweden	Switzerland	Australia	Luxembourg

10	Australia	United States	Switzerland	Australia	Sweden	Israel	Ireland
----	-----------	---------------	-------------	-----------	--------	--------	---------

- 행복 점수를 기준으로 Finland, Norway, Denmark, Iceland, Switzerland, Netherlands, Canada, New Zealand, Sweden, Australia가 상위 10개 국가(지역)로 선정되었다.

### 2-2-3. 2019년

#### 1) 각 변수 별 상위 10개 국가(지역)

<표 15> 2019년 각 변수 별 상위 10개 국가(지역)

rank	y(score)	gdp	support	life	freedom	generosity	corruption
1	Finland	Qatar	Iceland	Singapore	Uzbekistan	Myanmar	Singapore
2	Denmark	Luxembourg	Finland	Hong Kong	Cambodia	Indonesia	Rwanda
3	Norway	Singapore	Norway	Japan	Norway	Haiti	Denmark
4	Iceland	United Arab Emirates	Denmark	Spain	United Arab Emirates	Malta	Finland
5	Netherlands	Kuwait	New Zealand	Switzerland	Finland	Kenya	New Zealand
6	Switzerland	Ireland	Ireland	France	Denmark	Bhutan	Sweden
7	Sweden	Norway	Australia	Cyprus	Iceland	Thailand	Switzerland
8	New Zealand	Switzerland	United Kingdom	Northern Cyprus	New Zealand	Iceland	Norway
9	Canada	Hong Kong	Turkmenistan	Canada	Canada	United Kingdom	Luxembourg
10	Austria	United States	Mongolia	Italy	Sweden	Australia	Ireland

- 행복 점수를 기준으로 Finland, Denmark, Norway, Iceland, Netherlands, Switzerland, Sweden, New Zealand, Canada, Austria가 상위 10개 국가(지역)로 선정되었다.

- 높은 행복 점수를 받은 상위 국가들은 사회적 지원(support)과 자유(freedom)에서 상위권을 차지하였다. 탄탄한 사회 안전망을 가진 복지국가로 유명한 북유럽 국가들이 세계에서 행복한 국가들이라 할 수 있다.
- 표 12, 표 13, 표 14에 색칠되어 있는 국가는 상위 3개국으로, 종합적으로 보았을 때 gdp, life, generosity기준 상위 10개국에 대부분 포함되지 않는 것으로 판단된다.

### 2-3. 한국의 행복 점수

<표 16> 한국의 행복 점수 순위

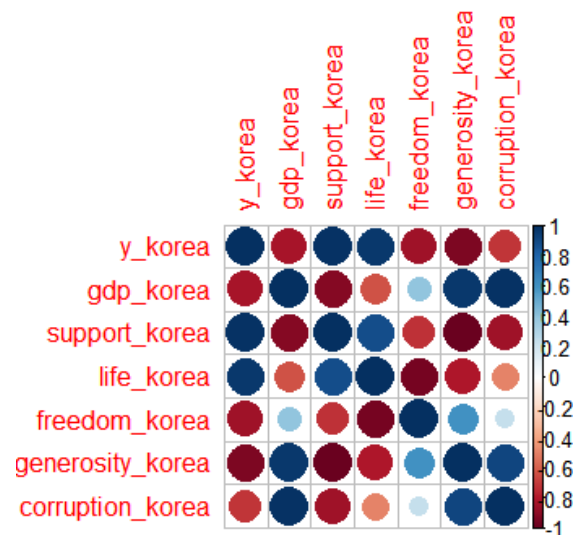
연도	y(score)	gdp	support	life	freedom	generosity	corruption
2017년	55	28	108	4	127	92	108
2018년	57	29	95	5	140	77	117
2019년	54	28	92	11	145	83	106

- 표 15는 한국의 행복 점수와 행복 점수의 기준 변수들에서의 순위를 나타낸 표이다.
- 2017년 한국의 행복 점수 순위는 155개국 중 55위를 차지하였다. 기대 수명(life)과 1인당 국내총생산(gdp)는 전체 국가 중 상위에 속하지만, 나머지 기준에서 하위권을 차지하고 있다. 특히 기대 수명(life)이 4위인 것과 달리 자유(freedom)가 127위로 현저히 낮은 순위를 차지하고 있다.
- 2018년 한국의 행복점수 순위는 156개국 중 57위로 2017년에 비해 순위가 두 계단 내려갔다. 사회적 지원(support)과 관용(generosity) 기준에서 순위가 상승한 반면 2018년에도 자유(freedom)에서 하위권을 차지하고 있으며 2017년보다 순위가 더 내려갔다.
- 2019년 한국의 행복점수 순위는 156개국 중 54위로 3개년 중 가장 높은 순위를 차지하였다. 1인당 국내총생산(gdp), 사회적 지원(support), 부정부패 인식(corruption) 기준에서 순위가 상승하였지만, 2년 동안 최상위권을 차지하고 있던 기대 수명(life) 순위가 떨어졌다. 또한 자유(freedom) 기준에서 순위가 더 떨어져 역시 하위권을 차지하고 있다.
- 데이터를 종합적으로 보아, 한국은 기대 수명(life)와 1인당 국내총생산(gdp) 기준에서 전체 국가 중 상위권을 차지하고 있지만 자유(freedom)과 부정부패 인식(corruption) 기준에서 순위가 매우 낮은 것을 알 수 있다. 이를 보아 자유(freedom)와 부정부패 인식(corruption)이 한국이 행복 점수 순위에 안 좋은 영향을 끼치는 것으로 판단된다.
- 한국의 행복 점수와 기준이 되는 변수들 간의 상관계수를 구했을 때, 사회적 자원(support), 기

대 수명(life)은 한국 행복 점수와 양의 상관관계를 갖는 반면 나머지 변수들은 음의 상관관계를 갖는다는 것을 알 수 있었다.

- 그림 30은 상관행렬 히트맵으로, y\_korea를 기준으로 한 첫번째 행만 살펴보면 된다. 사회적 자원(support), 기대 수명(life)은 히트맵에서 파란색 계열로 양의 상관관계인 반면 나머지 변수들은 빨간색 계열로 음의 상관관계를 나타낸다.

<그림 30> 한국 상관행렬 히트맵



### Ⅲ. 결론

#### 1. 기대효과

- OECD 회원국 중 자살률 1위인 한국의 행복 지수 순위가 평균 55위로 행복 지수 순위가 높지 않는다는 것을 알 수 있다. 이를 통해 자살률과 행복 지수와 관련되어 있는지 조사를 더 할 수 있지 않을까 기대해본다.

#### 2. 한계점

- 행복 점수를 계산할 때 국가별로 약 1000명에게 조사하는 것으로 알고 있는데, 표본의 수가 적다고 생각한다.

- 자유, 관용, 부정부패 인식 등과 같은 기준이 행복 점수를 계산한 결과가 객관성이 부족하다고 생각한다.

# R코드

## 1. 데이터 불러오기

```
# 2019 년
data1 <- read.csv("C:/Users/dldms/OneDrive/바탕 화면/회귀진단 레포트/Data/2019.csv")
str(data1)

# 2018 년
data2 <- read.csv("C:/Users/dldms/OneDrive/바탕 화면/회귀진단 레포트/Data/2018.csv", header=T)
str(data2)

data2$Perceptions.of.corruption[20] <- 0.112 # 결측치를 평균으로 대체
data2$Perceptions.of.corruption <- as.numeric(data2$Perceptions.of.corruption)

# 2017 년
data3 <- read.csv("C:/Users/dldms/OneDrive/바탕 화면/회귀진단 레포트/Data/2017.csv", header=T)
str(data3)

data1 <- cbind(data1, year=rep(0, 156)) # 2019
data2 <- cbind(data2, year=rep(1, 156)) # 2018
data3 <- cbind(data3, year=rep(2, 155)) # 2017

# 2017, 2018, 2019 년 합친 데이터
y <- c(data1$Score, data2$Score, data3$Happiness.Score)
gdp <- c(data1$GDP.per.capita, data2$GDP.per.capita, data3$Economy..GDP.per.Capita.)
support <- c(data1$Social.support, data2$Social.support, data3$Family)
life <- c(data1$Healthy.life.expectancy, data2$Healthy.life.expectancy, data3$Health..Life.Expectancy.)
freedom <- c(data1$Freedom.to.make.life.choices, data2$Freedom.to.make.life.choices, data3$Freedom)
generosity <- c(data1$Generosity, data2$Generosity, data3$Generosity)
corruption <- c(data1$Perceptions.of.corruption, data2$Perceptions.of.corruption, data3$Trust..Government.Corruption.)

year <- c(data1$year, data2$year, data3$year)
year <- as.factor(year)

data <- data.frame(y, gdp, support, life, freedom, generosity, corruption, year)

# 2017 년
y3 <- data3$Happiness.Score
gdp3 <- data3$Economy..GDP.per.Capita.
support3 <- data3$Family
life3 <- data3$Health..Life.Expectancy.
freedom3 <- data3$Freedom
generosity3 <- data3$Generosity
corruption3 <- data3$Trust..Government.Corruption.
region3 <- data3$Country
rank3 <- data3$Happiness.Rank

# 2018 년
y2 <- data2$Score
gdp2 <- data2$GDP.per.capita
support2 <- data2$Social.support
life2 <- data2$Healthy.life.expectancy
freedom2 <- data2$Freedom.to.make.life.choices
generosity2 <- data2$Generosity
corruption2 <- data2$Perceptions.of.corruption
region2 <- data2$Country.or.region
rank2 <- data2$Overall.rank

# 2019 년
y1 <- data1$Score
gdp1 <- data1$GDP.per.capita
support1 <- data1$Social.support
life1 <- data1$Healthy.life.expectancy
freedom1 <- data1$Freedom.to.make.life.choices
generosity1 <- data1$Generosity
corruption1 <- data1$Perceptions.of.corruption
region1 <- data1$Country.or.region
rank1 <- data1$Overall.rank
```

## 2. 요약통계량

```
# 2017 년
mean(y3);quantile(y3)

mean(gdp3);quantile(gdp3)

mean(support3);quantile(gdp3)

mean(life3);quantile(life3)

mean(freedom3);quantile(freedom3)

mean(generosity3);quantile(generosity3)

mean(corruption3);quantile(corruption3)

# 2018 년
mean(y2);quantile(y2)

mean(gdp2);quantile(gdp2)

mean(support2);quantile(support2)

mean(life2);quantile(life2)

mean(freedom2);quantile(freedom2)

mean(generosity2);quantile(generosity2)

mean(corruption2);quantile(corruption2)

# 2018 년
mean(y1);quantile(y1)

mean(gdp1);quantile(gdp1)

mean(support1);quantile(gdp1)

mean(life1);quantile(life1)

mean(freedom1);quantile(freedom1)

mean(generosity1);quantile(generosity1)

mean(corruption1);quantile(corruption1)
```

## 3. 변수 그림(히스토그램)

```
## 2017 년 데이터 히스토그램

hist(y3,xlab="Score",main="Happiness Score")

hist(gdp3,xlab="GDP",main="GDP")

hist(data3$Family,xlab="Social support",main="Social support")

hist(life3,xlab="Healthy life expectancy",main="Healthy life expectancy")

hist(freedom3,xlab="Freedom to make life choices",main="Freedom to make life choices")

hist(generosity3,xlab="Generosity",main="Generosity")

hist(corruption3,xlab="Perceptions of corruption",main="Perceptions of corruption")

# 로그 변환
generosity3_log <- log(generosity3)
generosity3_log[is.infinite(generosity3_log)] <- 0
corruption3_log <- log(corruption3)
corruption3_log[is.infinite(corruption3_log)] <- 0
```



```

hist(generosity3,xlab="Generosity",main="Generosity")
hist(generosity3_log,xlab="log Generosity",main="Generosity")
hist(corruption3,xlab="Perceptions of corruption",main="Perceptions of corruption")
hist(corruption3_log,xlab="log Perceptions of corruption",main="Perceptions of corruption")

## 2018 년 데이터 히스토그램
hist(y2,xlab="Score",main="Happiness Score")

hist(gdp2,xlab="GDP",main="GDP")

hist(support2,xlab="Social support",main="Social support")

hist(life2,xlab="Healthy life expectancy",main="Healthy life expectancy")

hist(freedom2,xlab="Freedom to make life choices",main="Freedom to make life choices")

hist(generosity2,xlab="Generosity",main="Generosity")

hist(corruption2,xlab="Perceptions of corruption",main="Perceptions of corruption")

# 로그 변환
generosity2_log <- log(generosity2)
generosity2_log[is.infinite(generosity2_log)] <- 0
corruption2_log <- log(corruption2)
corruption2_log[is.infinite(corruption2_log)] <- 0

hist(generosity2,xlab="Generosity",main="Generosity")

hist(generosity2_log,xlab="log Generosity",main="Generosity")

hist(corruption2,xlab="Perceptions of corruption",main="Perceptions of corruption")

hist(corruption2_log,xlab="log Perceptions of corruption",main="Perceptions of corruption")

## 2019 년 데이터 히스토그램
hist(y1,xlab="Score",main="Happiness Score")

hist(gdp1,xlab="GDP",main="GDP")

hist(support1,xlab="Social support",main="Social support")

hist(life1,xlab="Healthy life expectancy",main="Healthy life expectancy")

hist(freedom1,xlab="Freedom to make life choices",main="Freedom to make life choices")

hist(generosity1,xlab="Generosity",main="Generosity")

hist(corruption1,xlab="Perceptions of corruption",main="Perceptions of corruption")

# 로그 변환
generosity1_log <- log(generosity1)
generosity1_log[is.infinite(generosity1_log)] <- 0
corruption1_log <- log(corruption1)
corruption1_log[is.infinite(corruption1_log)] <- 0

hist(generosity1,xlab="Generosity",main="Generosity")

hist(generosity1_log,xlab="log Generosity",main="Generosity")

hist(corruption1,xlab="Perceptions of corruption",main="Perceptions of corruption")

hist(corruption1_log,xlab="log Perceptions of corruption",main="Perceptions of corruption")

```

```
## 2017, 2018, 2019 년 합친 데이터 히스토그램
hist(y,xlab="Score",main="Happiness Score")

hist(gdp,xlab="GDP",main="GDP")

hist(support,xlab="Social support",main="Social support")

hist(life,xlab="Healthy life expectancy",main="Healthy life expectancy")

hist(freedom,xlab="Freedom to make life choices",main="Freedom to make life choices")

hist(generosity,xlab="Generosity",main="Generosity")

hist(corruption,xlab="Perceptions of corruption",main="Perceptions of corruption")

# 로그 변환
generosity_log <- log(generosity)
generosity_log[is.infinite(generosity_log)] <- 0
corruption_log <- log(corruption)
corruption_log[is.infinite(corruption_log)] <- 0

hist(generosity,xlab="Generosity",main="Generosity")

hist(generosity_log,xlab="log Generosity",main="Generosity")

hist(corruption,xlab="Perceptions of corruption",main="Perceptions of corruption")

hist(corruption_log,xlab="log Perceptions of corruption",main="Perceptions of corruption")
```

#### 4. 연도 별 행복점수(y)와 독립변수 간에 상관관계 파악

```
## 2019 년
cor.test(y1,gdp1)

cor.test(y1,support1)

cor.test(y1,life1)

cor.test(y1,freedom1)

cor.test(y1,generosity1)

cor.test(y1,corruption1)

data1_part_log <- data.frame(y1,gdp1,support1,life1,freedom1,generosity1_log,corruption1_log)
head(data1_part_log)

library(corrplot)

cor1_log <- cor(data1_part_log)
round(cor1_log,3)

corrplot(cor1_log) # 상관행렬 히트맵

## 2018 년
cor.test(y2,gdp2)

cor.test(y2,support2)

cor.test(y2,life2)

cor.test(y2,freedom2)

cor.test(y2,generosity2)

cor.test(y2,corruption2)
```

```

data2_part_log <- data.frame(y2,gdp2,support2,life2,freedom2,generosity2_log,corruption2_log)
head(data2_part_log)

cor2_log <- cor(data2_part_log)
round(cor2_log,3)

corrplot(cor2_log) # 상관행렬 히트맵

## 2017 년
cor.test(y3,gdp3)

cor.test(y3,support3)

cor.test(y3,life3)

cor.test(y3,freedom3)

cor.test(y3,generosity3)

cor.test(y3,corruption3)

data3_part_log <- data.frame(y3,gdp3,support3,life3,freedom3,generosity3_log,corruption3_log)
head(data3_part_log)

cor3_log <- cor(data3_part_log)
round(cor3_log,3)

corrplot(cor3_log) # 상관행렬 히트맵

```

## 5. 모형 선택

```

fit1 <- lm(y~gdp+support+life+freedom+generosity+corruption+year) # 모형 1
fit2 <- lm(y~gdp+support+life+freedom+generosity+corruption_log+year) # 모형 2
fit3 <- lm(y~gdp+support+life+freedom+generosity_log+corruption+year) # 모형 3
fit4 <- lm(y~gdp+support+life+freedom+generosity_log+corruption_log+year) # 모형 4

car::vif(fit1) # 다중공선성 판단

summary(fit1)

car::vif(fit2) # 다중공선성 판단

summary(fit2)

car::vif(fit3) # 다중공선성 판단

summary(fit3)

car::vif(fit4) # 다중공선성 판단

summary(fit4)

# 모형 진단
plot(fit1)

plot(fit2)

plot(fit3)

```

```
plot(fit4)
```

## 6. 특이값, 영향력이 큰 관측값 판단

```
# 독립변수 기준으로 특이값
hatvalues(fit4)[2*mean(hatvalues(fit4))]

# 종속변수 기준으로 특이값
rstudent(fit4)[rstudent(fit4)>qt(0.05/2,467-7-2,lower.tail=FALSE)]

# 영향력이 큰 관측값
cooks.distance(fit4)[cooks.distance(fit4)>qf(0.5,7+1,467-7-1,lower.tail=FALSE)]
```

## 7. 변수 선택 : 단계별 회귀 방법

```
# 변수 선택 : 단계별 회귀방법
fit_null <- lm(y~1)
step(fit_null,scope=list(lower=~1,upper=~gdp+support+life+freedom+generosity_log+corruption_log),direction="both")
```

## 8. 2017 년 변수별 상위 10 개 국가(지역)

```
# 2017 년
gdp3_sort <- sort(gdp3,decreasing = TRUE)[1:10]
gdp3_region <- c()

for(i in 1:10){
  for(j in 1:155){
    if(gdp3_sort[i]==gdp3[j]){
      gdp3_region <- c(gdp3_region,region3[j])
    }
  }
}

support3_sort <- sort(support3,decreasing = TRUE)[1:10]
support3_region <- c()

for(i in 1:10){
  for(j in 1:155){
    if(support3_sort[i]==support3[j]){
      support3_region <- c(support3_region,region3[j])
    }
  }
}

life3_sort <- sort(life3,decreasing = TRUE)[1:10]
life3_region <- c()

for(i in 1:10){
  for(j in 1:155){
    if(life3_sort[i]==life3[j]){
      life3_region <- c(life3_region,region3[j])
    }
  }
}

freedom3_sort <- sort(freedom3,decreasing = TRUE)[1:10]
freedom3_region <- c()

for(i in 1:10){
  for(j in 1:155){
    if(freedom3_sort[i]==freedom3[j]){
      freedom3_region <- c(freedom3_region,region3[j])
    }
  }
}

generosity3_sort <- sort(generosity3,decreasing = TRUE)[1:10]
generosity3_region <- c()

for(i in 1:10){
  for(j in 1:155){
    if(generosity3_sort[i]==generosity3[j]){
```

```

        generosity3_region <- c(generosity3_region,region3[j])
    }
}

corruption3_sort <- sort(corruption3,decreasing = TRUE)[1:10]
corruption3_region <- c()

for(i in 1:10){
    for(j in 1:155){
        if(corruption3_sort[i]==corruption3[j]){
            corruption3_region <- c(corruption3_region,region3[j])
        }
    }
}

cbind(gdp3_region,support3_region,life3_region,freedom3_region,generosity3_region,corruption3_region,region3[1:10])

```

## 9. 2018 년 변수별 상위 10 개 국가(지역)

```

# 2018 년
gdp2_sort <- sort(gdp2,decreasing = TRUE)[1:10]
gdp2_region <- c()

for(i in 1:10){
    for(j in 1:156){
        if(gdp2_sort[i]==gdp2[j]){
            gdp2_region <- c(gdp2_region,region2[j])
        }
    }
}

support2_sort <- sort(support2,decreasing = TRUE)[1:10]
support2_region <- c()

for(i in c(1:10)){
    for(j in 1:156){
        if(support2_sort[i]==support2[j]){
            support2_region <- c(support2_region,region2[j])
        }
    }
}

life2_sort <- sort(life2,decreasing = TRUE)[1:10]
life2_region <- c()

for(i in 1:10){
    for(j in 1:156){
        if(life2_sort[i]==life2[j]){
            life2_region <- c(life2_region,region2[j])
        }
    }
}

freedom2_sort <- sort(freedom2,decreasing = TRUE)[1:10]
freedom2_region <- c()

for(i in 1:10){
    for(j in 1:156){
        if(freedom2_sort[i]==freedom2[j]){
            freedom2_region <- c(freedom2_region,region2[j])
        }
    }
}

generosity2_sort <- sort(generosity2,decreasing = TRUE)[1:10]
generosity2_region <- c()

for(i in c(1,2,3,4,6,7,9,10)){ # 중복 제외
    for(j in 1:156){
        if(generosity2_sort[i]==generosity2[j]){
            generosity2_region <- c(generosity2_region,region2[j])
        }
    }
}

corruption2_sort <- sort(corruption2,decreasing = TRUE)[1:10]
corruption2_region <- c()

```

```

for(i in 1:10){
  for(j in 1:156){
    if(corruption2_sort[i]==corruption2[j]){
      corruption2_region <- c(corruption2_region,region2[j])
    }
  }
}

cbind(gdp2_region,support2_region,life2_region,freedom2_region,generosity2_region,corruption2_region,region2[1:10])

```

## 10. 2019 년 변수별 상위 10 개 국가(지역)

```

# 2019 년
gdp1_sort <- sort(gdp1,decreasing = TRUE)[1:10]
gdp1_region <- c()

for(i in 1:10){
  for(j in 1:156){
    if(gdp1_sort[i]==gdp1[j]){
      gdp1_region <- c(gdp1_region,region1[j])
    }
  }
}

support1_sort <- sort(support1,decreasing = TRUE)[1:10]
support1_region <- c()

for(i in c(1:8,10)){ # 중복 제외
  for(j in 1:156){
    if(support1_sort[i]==support1[j]){
      support1_region <- c(support1_region,region1[j])
    }
  }
}

life1_sort <- sort(life1,decreasing = TRUE)[1:10]
life1_region <- c()

for(i in c(1:7,9)){ # 중복 제외
  for(j in 1:156){
    if(life1_sort[i]==life1[j]){
      life1_region <- c(life1_region,region1[j])
    }
  }
}

freedom1_sort <- sort(freedom1,decreasing = TRUE)[1:10]
freedom1_region <- c()

for(i in 1:10){
  for(j in 1:156){
    if(freedom1_sort[i]==freedom1[j]){
      freedom1_region <- c(freedom1_region,region1[j])
    }
  }
}

generosity1_sort <- sort(generosity1,decreasing = TRUE)[1:10]
generosity1_region <- c()

for(i in 1:10){
  for(j in 1:156){
    if(generosity1_sort[i]==generosity1[j]){
      generosity1_region <- c(generosity1_region,region1[j])
    }
  }
}

corruption1_sort <- sort(corruption1,decreasing = TRUE)[1:10]
corruption1_region <- c()

for(i in 1:10){
  for(j in 1:156){
    if(corruption1_sort[i]==corruption1[j]){
      corruption1_region <- c(corruption1_region,region1[j])
    }
  }
}

```

```

    }
  }
  cbind(gdp1_region,support1_region,life1_region,freedom1_region,generosity1_region,corruption1_region,region1[1:10])

```

#### 11. 2017 년 변수별로 한국이 차지한 순위

```

# 2017 년
k3 <- rank3[region3=="South Korea"]
rank(-gdp3)[k3]

rank(-support3)[k3]

rank(-life3)[k3]

rank(-freedom3)[k3]

rank(-generosity3)[k3]

rank(-corruption3)[k3]

cbind(k3,gdp3=rank(-gdp3)[k3],support3=rank(-support3)[k3],life3=rank(-life3)[k3],freedom3=rank(-freedom3)[k3],generosity3=rank(-generosity3)[k3],corruption3=rank(-corruption3)[k3])

```

#### 12. 2018 년 변수별로 한국이 차지한 순위

```

# 2018 년
k2 <- rank2[region2=="South Korea"]
rank(-gdp2)[k2]

rank(-support2)[k2]

rank(-life2)[k2]

rank(-freedom2)[k2]

rank(-generosity2)[k2]

rank(-corruption2)[k2]

cbind(k2,gdp2=rank(-gdp2)[k2],support2=rank(-support2)[k2],life2=rank(-life2)[k2],freedom2=rank(-freedom2)[k2],generosity2=rank(-generosity2)[k2],corruption2=rank(-corruption2)[k2])

```

#### 13. 2019 년 변수별로 한국이 차지한 순위

```

# 2019 년
k1 <- rank1[region1=="South Korea"]
rank(-gdp1)[k1]

rank(-support1)[k1]

rank(-life1)[k1]

rank(-freedom1)[k1]

rank(-generosity1)[k1]

rank(-corruption1)[k1]

cbind(k1,gdp1=rank(-gdp1)[k1],support1=rank(-support1)[k1],life1=rank(-life1)[k1],freedom1=rank(-freedom1)[k1],generosity1=rank(-generosity1)[k1],corruption1=rank(-corruption1)[k1])

```

#### 14. 2017, 2018, 2019 년 변수별로 한국이 차지한 순위를 합친 데이터에서 행복 점수(y)와 변수별로 상관관계 비교

```
y_korea <- c(y1[k1],y2[k2],y3[k3])
gdp_korea <- c(gdp1[k1],gdp2[k2],gdp3[k3])
support_korea <- c(support1[k1],support2[k2],support3[k3])
life_korea <- c(life1[k1],life2[k2],life3[k3])
freedom_korea <- c(freedom1[k1],freedom2[k2],freedom3[k3])
generosity_korea <- c(generosity1[k1],generosity2[k2],generosity3[k3])
corruption_korea <- c(corruption1[k1],corruption2[k2],corruption3[k3])

data_korea <- data.frame(y_korea,gdp_korea,support_korea,life_korea,freedom_korea,generosity_korea,corruption_korea)

# 상관계수
cor(y_korea,gdp_korea)

cor(y_korea,support_korea)

cor(y_korea,life_korea)

cor(y_korea,freedom_korea)

cor(y_korea,generosity_korea)

cor(y_korea,corruption_korea)

round(cor(data_korea),3)

corrplot(round(cor(data_korea),3)) # 상관행렬 비트맵
```

## 참고문헌

- [1] 이우리(2012). <회귀분석 입문 – 이해와 SAS 활용>. 탐진
- [2] 서민구(2014). <R을 이용한 데이터 처리 & 분석 실무>. 길벗.  
<https://thebook.io/006723/ch08/02/06/>
- [3] 서현우. (2020.04.28). [초점 | 2020행복보고서] 한국 세계행복지수 7계단 하락!. 월간산,  
<http://san.chosun.com/m/svc/article.html?contid=2020042203597>
- [4] 손영하. (2020.09.22). '자살 공화국' 오명 언제 벗을까... "올해도 OECD 압도적 1위". 한국일보,  
<https://www.hankookilbo.com/News/Read/A2020092210030004232>
- [5] Jin-Hoon An. 통계 분석 기법을 이용한 가설 검정. <https://rpubs.com/jmhome/StatisticalAnalysis>
- [6] Keon-Woong Moon. R강의6. 회귀진단과 모형의 선택. [https://rstudio-pubs-static.s3.amazonaws.com/190997\\_40fa09db8e344b19b14a687ea5de914b.html](https://rstudio-pubs-static.s3.amazonaws.com/190997_40fa09db8e344b19b14a687ea5de914b.html)
- [7] mathurinache. <https://www.kaggle.com/mathurinache/world-happiness-report>
- [8] Sustainable Development Solutions Network. <https://www.kaggle.com/unsdsn/world-happiness>