

ch2. 비건 식품이 환경 및 동물 보호에 미치는 영향 분석

4 조

2022. 5. 13

먼저, 식품이 환경에 미치는 영향을 데이터 시각화를 통해 확인하고, 환경보호를 위해서 어떤 식품들의 섭취를 줄이는 것이 좋은지 알아보겠습니다.

그리고 온실가스 증가가 동물보호에 영향을 주는지 연관분석 및 회귀분석을 통해 검증하고자 합니다.

비건 스타일의 식생활을 하면, (비건 아닌 식습관대비) 상대적으로 더 많은 온실가스 배출량을 줄일 수 있고, 온실가스 배출량 증가가 줄면, 멸종위기동물 수 감소에 영향을 줄 것이라는 가설을 검정해보겠습니다.

이를 통해 비건이 환경과 동물 보호에 긍정적인 영향을 주는지 살펴보겠습니다.

```
# 식품별 온실가스 배출량 및 음식물쓰레기량 raw data
library(readxl)
c02data <- read_excel("c02data.xls")
str(c02data)

## tibble [43 x 5] (S3: tbl_df/tbl/data.frame)
## $ Category      : chr [1:43] "Wine" "Vegetable Oil" "Vegetable Oil"
## $ Product       : chr [1:43] "Wine" "PalmOil" "OliveOil" "SoybeanOil"
## $ GHGmissionperkg2013: num [1:43] 1.59 7.19 5.09 3.87 3.53 3.52 3.17 1.7
## $ GHGmissionperkg2007: num [1:43] 1.57 6.97 5.04 3.83 3.54 3.53 3.21 1.7
## $ FoodWasteton    : num [1:43] 26013000 16691000 2997000 24148000
## $ FoodWasteton    : num [1:43] 9554000 ...
```

1. 식품별 온실가스 배출량 시각화

a. 식품별 발생하는 온실가스 배출량 수치 확인

- 식품 카테고리별 온실가스 배출량 평균 집계
 - > 인사이트 : ****고기 > 생선 > 유제품 > 커피 > 초콜릿**** 순으로 많은 온실가스 발생함.
- 기술통계량
 - > 식품 카테고리 총 12 개,
****온실가스 평균 배출량(kg 당) 6.34, 최소량 0.6, 최대량 30.7****
- 식품 카테고리별 온실가스 배출 비중
 - > ****고기(40.2%), 생선(14.8%), 유제품(11.2%)을 먹지 않으면,****
개인이 배출하는 온실가스 배출량의 **약 66%를 감소시키는 효과 발생.**

#식품 카테고리별 평균 집계

```
library(dplyr)
c02data1 <- c02data %>% select(Category, GHGemissionperkg2013) %>%
  group_by(Category) %>%
  summarise(mean_GHG_2013 = round(mean(GHGemissionperkg2013), digits = 1)) %>%
  arrange(desc(mean_GHG_2013))
```

#식품 카테고리 범주형 변환

```
c02data1$Category <- factor(c02data1$Category,
                             levels = c("Meat", "Fish", "Dairy", "Coffe",
                                           "Chocolate", "Vegetable Oil", "Grain",
                                           "Soy Food", "Wine", "Vegetable", "Bean",
                                           "Fruit"))
```

#표 만들기

```
library(knitr)
kable(head(c02data1))
```

Category	mean_GHG_2013
Meat	30.6
Fish	11.3

Category	mean_GHG_2013
Dairy	8.5
Coffe	8.2
Chocolate	5.0
Vegetable Oil	4.6

#기술통계량 표 확인

```
library(psych)
kable(describe(c02data1$mean_GHG_2013))
```

	va				med	trim		mi	m	ran		kurto	
	rs	n	mean	sd	ian	med	mad	n	ax	ge	skew	sis	se
X	1	1	6.341	8.414	3.3	4.48	3.48	0.	30	29.	1.905	2.834	2.429
1		2	667	539			411	7	.6	9	648	058	068

#식품별 온실가스 발생 비중 컬럼 추가, 컬럼명 변경

```
data_prop <- c02data1 %>%
  mutate(paste0(prop =
    round((mean_GHG_2013/sum(mean_GHG_2013)*100),2),"%"))
names(data_prop) <- c("Category", "GreenhouseGas", "Percent")
```

#표 만들기

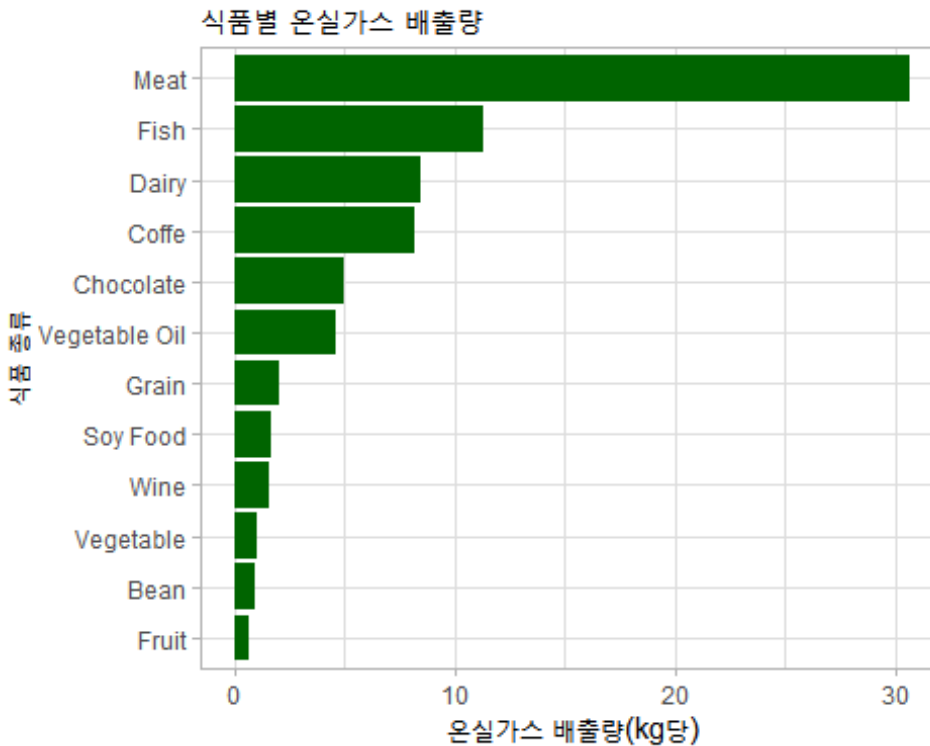
```
kable(data_prop,
  align=c("l","r","r"))
```

Category	GreenhouseGas	Percent
Meat	30.6	40.21%
Fish	11.3	14.85%
Dairy	8.5	11.17%
Coffe	8.2	10.78%
Chocolate	5.0	6.57%
Vegetable Oil	4.6	6.04%
Grain	2.0	2.63%
Soy Food	1.7	2.23%
Wine	1.6	2.1%
Vegetable	1.0	1.31%
Bean	0.9	1.18%
Fruit	0.7	0.92%

b. 그래프 시각화

```
library(ggplot2)
ggplot(c02data1, aes(x=reorder(Category, mean_GHG_2013), y=mean_GHG_2013)) +
```

```
geom_col(fill="darkgreen") +
  theme_light() + labs(x= "식품 종류", y= "온실가스 배출량(kg 당)", title= "식품
별 온실가스 배출량") + coord_flip()
```



```
#식품별 온실가스 발생 비중 원형 차트로 시각화
library(readr)
data_prop_numeric <- parse_number(data_prop$Percent) #텍스트를 숫자로 변환
data_prop2 <- data_prop
data_prop2$Percent <- data_prop_numeric

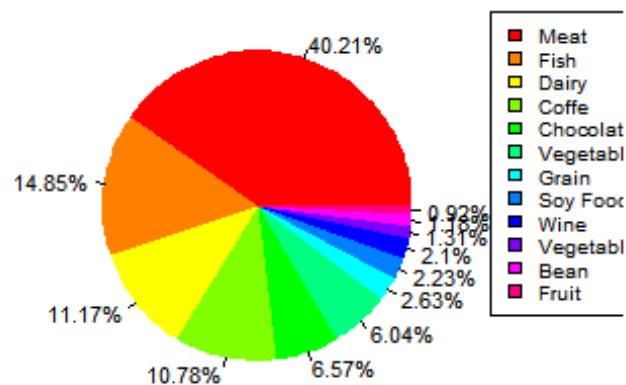
#반응형 파이차트 (R Markdown html 용)
# library(plotly) #반응형 그래프 패키지 로드
# library(dplyr)
# p <- plot_ly(data_prop2, labels = ~Category, values = ~Percent, textinfo='l
abel+percent',type = 'pie',
# maker = list(colors = ~colors)) %>%
# layout(title = "식품별 온실가스 배출 비중",
# xaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels =
FALSE),
# yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels =
FALSE))
```

```
# p
```

```
# 일반 파이차트 (R Markdown Word 용)
```

```
pie(data_prop2$Percent, labels = paste0(data_prop2$Percent, "%"),  
     main = "식품별 온실가스 배출 비중", col=rainbow(12), cex.main=1.5, col.main  
     ="black", cex=0.7, radius=0.8, border=NA)  
legend(1.2, 1, data_prop2$Category, fill=rainbow(12), cex = 0.6)
```

식품별 온실가스 배출 비중



=====

2. CO2 증가에 따른 멸종위기 동물 수 회귀분석

```
library(dplyr)  
library(knitr)
```

```
## 연도별 글로벌 CO2 증가 raw data 불러옴
```

```
c02inc <- read.csv("c02increase.csv")
```

```
# 연도 날짜 포맷으로 변경
```

```
c02inc$Year <- format(as.Date(c02inc$Year, "%Y"), "%Y")  
kable(head(c02inc))
```

Year	Annual.Increase	Uncertainty
1959	0.94	0.11
1960	0.54	0.11
1961	0.95	0.11
1962	0.64	0.11
1963	0.71	0.11
1964	0.28	0.11

##연도별 멸종위기동물 raw data 불러옴

```
mammal <- read.csv("redlistindex.csv")
mammal <- mammal[,1:2]
mammal$Year <- as.character(mammal$Year) #연도 문자형 변환
kable(head(mammal))
```

Year	Threatened.Species_Mammal
2021	1333
2020	1323
2019	1244
2018	1219
2017	1204
2016	1194

##데이터 결합 및 변수명 변경 (이너조인)

```
c02_red_mal <- inner_join(mammal, c02inc, by="Year")
c02_red_mal <- c02_red_mal[,1:3]
names(c02_red_mal) <- c("Year", "threatened_animals", "c02_increase")
kable(head(c02_red_mal))
```

Year	threatened_animals	c02_increase
2017	1204	1.89
2016	1194	2.99
2015	1197	3.02
2014	1199	2.18
2013	1143	2.01
2012	1139	2.61

1) 탐색적 자료분석 (EDA)

- 데이터 설명 : 시계열데이터로서, 연간 c02 증가량 및 연간 멸종위기 포유류 수를 나타낸다. - c02_increase : Annual CO2 mole fraction increase (ppm) -
threatened_animals : Numbers of threatened species by Mammals

- b. 기초통계량 확인 : - C02 증가량의 최저치는 1.160, 최대치는 3.020 이며, 중앙값은 2.010 이다. - 멸종위기동물의 최저수는 535, 최대수는 1204 이며, 중앙값은 1137 이다.

```
summary(c02_red_mal)
```

```
##      Year      threatened_animals  c02_increase
## Length:19      Min.   : 535      Min.   :1.160
## Class :character 1st Qu.:1098      1st Qu.:1.645
## Mode  :character Median :1137      Median :2.010
##                Mean   :1089      Mean   :2.041
##                3rd Qu.:1142      3rd Qu.:2.295
##                Max.   :1204      Max.   :3.020
```

- c. 그래프 시각화 : 두 그래프가 우상향으로 선형성을 보인다.

```
library(ggplot2)
```

```
#연도별 c02 증가 막대그래프
```

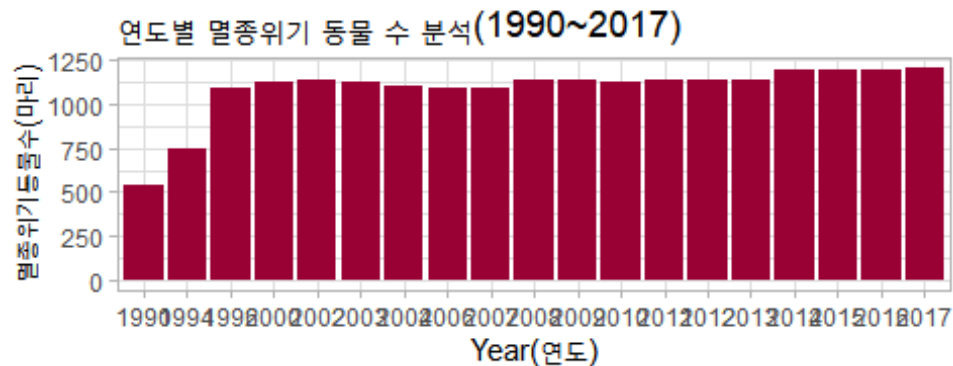
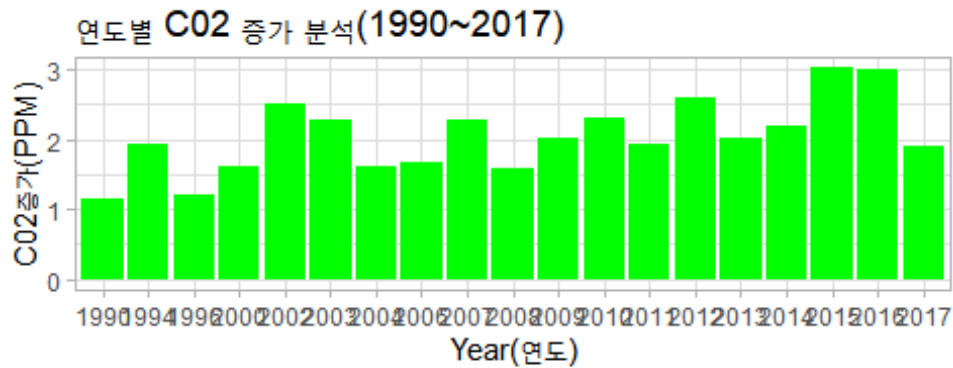
```
p1 <- ggplot() + geom_col(aes(x=Year, y=c02_increase), data=c02_red_mal, fill="green") +  
  theme_light() +  
  labs(x = "Year(연도)",  
       y = "C02 증가(PPM)",  
       title = "연도별 c02 증가 분석(1990~2017)")
```

```
#연도별 멸종위기종 수 막대그래프
```

```
p2 <- ggplot() + geom_col(aes(x=Year, y=threatened_animals), data=c02_red_mal,  
  fill="#990033") +  
  theme_light() +  
  labs(x = "Year(연도)",  
       y = "멸종위기동물수(마리)",  
       title = "연도별 멸종위기 동물 수 분석(1990~2017)")
```

```
#두 그래프 한페이지에 보기
```

```
library(gridExtra)  
grid.arrange(p1, p2, nrow = 2)
```



2) 상관분석 및 회귀분석

a. F 검정 (등분산검정) - 귀무가설 : c02 증가와 멸종위기동물수 모분산에 차이가 없다 -
대립가설 : c02 증가와 멸종위기동물수 모분산에 차이가 있다 -> $p\text{-value} < 2.2e-16$ 이므로,
귀무가설을 기각한다. -> 모분산에 차이가 있으므로, 등분산 가정에 만족하지 않는다고 볼 수 있다.

b. 상관분석 (연관관계) - 귀무가설 : 두변수간(c02 증가, 멸종위기동물수) 선형관계가 존재하지 않는다 - 대립가설 : 두변수간(c02 증가, 멸종위기동물수) 선형관계가 존재한다 -
> $p\text{-value} = 0.03982$ 이므로, 귀무가설을 기각하여 상관관계가 존재한다고 볼 수 있다. -> 상관계수는 0.4750949 로, 양의 상관관계를 갖는다.

c. 회귀분석 (인과관계) - 귀무가설 : C02 증가는 멸종위기동물 수에 영향을 주지 않는다 -
대립가설 : C02 증가는 멸종위기동물 수에 영향을 준다

- **회귀식** : $\text{threatened_animals}(y) = 152.45 * \text{c02_increase}(x) + 777.53$
-> **C02 가 1PPM 증가하면, 멸종위기동물이 929.98 마리 발생한다.**

- 회귀 계수 검정 : 회귀계수 t-통계량의 p-value 가 0.05 보다 작으면 통계적으로 유의
-> $pr\ 0.0398 < 0.05$ 이므로 통계적으로 유의함.
- ****회귀 모형 설명력 - 결정 계수 (R^2)**** : 전체 데이터를 회귀 모형이 얼마나 잘 설명하고 있는지 보여주는 지표 (회귀선의 정확도 평가)
(전체 제곱 합에서 회귀 제곱 합의 비율 $R^2 = SSR/SST = 1 - SSE/SST$)
Multiple R-squared: 0.2257, Adjusted R-squared: 0.1802
-> ****c02 증가가 멸종위기동물수의 변동을 22.57% 설명한다.****
- ****통계적 유의성 검정**** : 회귀 분석 결과로 산출되는 F-통계량의 p-값이 0.05 보다 작으면 회귀 모형은 통계적으로 유의
-> $p\text{-value}: 0.03982 < 0.05$ 이므로 통계적으로 유의함.
귀무가설 기각하고 대립가설 채택한다.

결론 : c02 증가는 멸종위기동물 수에 영향을 준다고 볼 수 있다.

```
var.test(c02_red_mal$threatened_animals, c02_red_mal$c02_increase)

##
## F test to compare two variances
##
## data: c02_red_mal$threatened_animals and c02_red_mal$c02_increase
## F = 102964, num df = 18, denom df = 18, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 39668.82 267252.80
## sample estimates:
## ratio of variances
## 102964.1

cor.test(c02_red_mal$threatened_animals, c02_red_mal$c02_increase)

##
## Pearson's product-moment correlation
##
## data: c02_red_mal$threatened_animals and c02_red_mal$c02_increase
## t = 2.2261, df = 17, p-value = 0.03982
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.02663271 0.76436082
## sample estimates:
## cor
## 0.4750949
```

```

model_lm_5 <- lm(threatened_animals ~ c02_increase, data = c02_red_mal)
summary(model_lm_5)

##
## Call:
## lm(formula = threatened_animals ~ c02_increase, data = c02_red_mal)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -419.37  -33.00   56.53   84.34  138.34
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    777.53     143.97   5.401 4.78e-05 ***
## c02_increase    152.45      68.48   2.226  0.0398 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 150.5 on 17 degrees of freedom
## Multiple R-squared:  0.2257, Adjusted R-squared:  0.1802
## F-statistic: 4.956 on 1 and 17 DF,  p-value: 0.03982

```

D. 그래프 확인

```

library(ggpubr)

#산점도 그래프로 x, y 상관관계 확인
ggscatter(c02_red_mal, x = "c02_increase", y = "threatened_animals",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "C02 증가 (PPM)", ylab = "멸종위기 동물 수 (마리)", title = "C
02 증가와 멸종위기 동물 수 상관관계")

```

C02 증가와 멸종위기 동물 수 상관관계

