

BIM503 Homework1

PMBA 9 20214044 윤진서

BIM503 Homework 1

분석에 사용할 데이터를 email.df 변수로 불러오고 구조를 확인합니다.

```
email.df = read.csv("res/L2_email_campaign.csv")
str(email.df)
```

```
## 'data.frame': 1000 obs. of 8 variables:
## $ id : int 1 2 3 4 5 6 7 8 9 10 ...
## $ respond : int 0 0 0 1 0 0 0 0 0 1 ...
## $ gender : int 1 1 1 1 1 1 1 1 1 1 ...
## $ age : int 37 41 37 36 36 33 44 43 35 55 ...
## $ numpurchase: int 1 2 1 2 1 2 1 2 2 8 ...
## $ totdol : int 59 118 60 497 64 61 62 50 53 4339 ...
## $ region : int 6 1 4 2 2 3 4 1 1 3 ...
## $ addresstype: chr "urban" "rural" "urban" "urban" ...
```

Question 1.

```
summary(email.df$age)
```

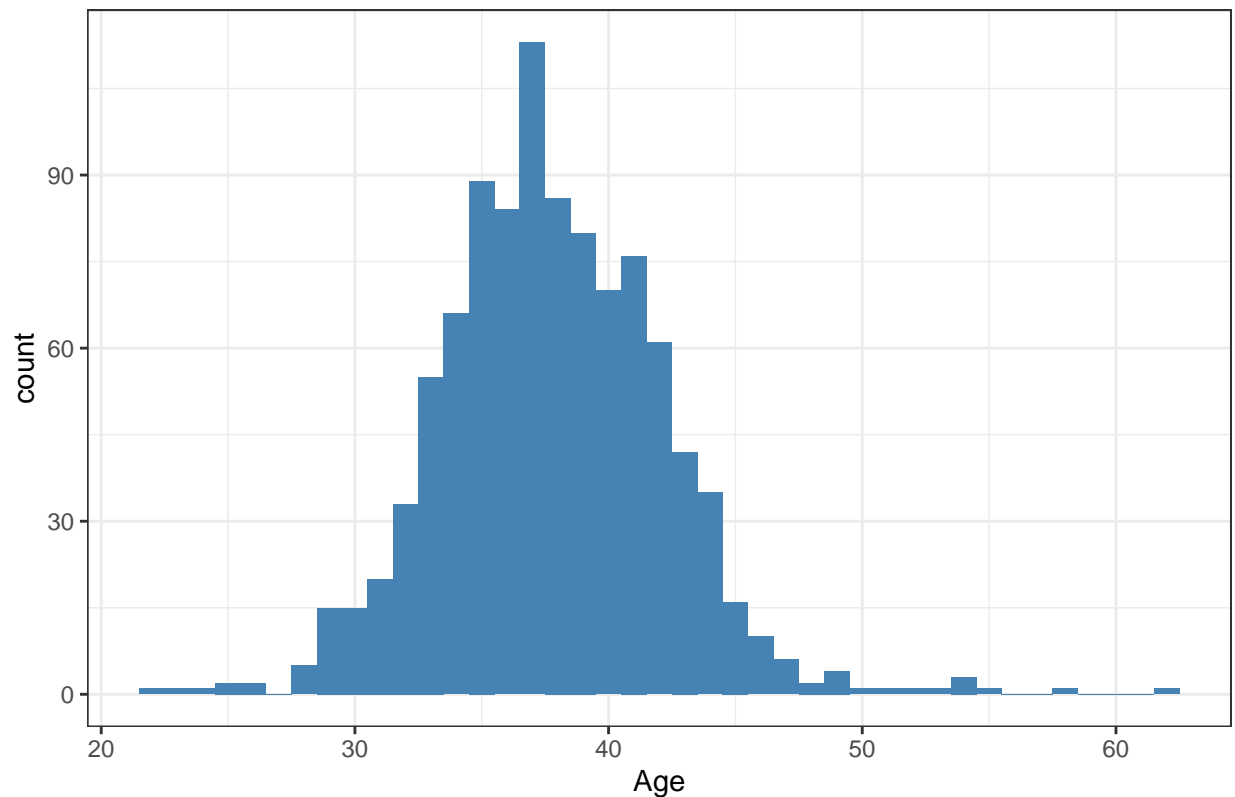
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 22.00   35.00   37.00   37.76   41.00   62.00
```

Question 2.

```
library(ggplot2)
library(ggthemes)

ggplot(email.df, aes(x=age)) +
  geom_histogram(fill="steelblue", binwidth=1) +
  xlab("Age") +
  ggtitle("Histogram for Age") +
  theme_bw()
```

Histogram for Age



Question 3.

```
t.test(email.df$age, mu=40, alternative = "less")
```

```
##
## One Sample t-test
##
## data: email.df$age
## t = -16.049, df = 999, p-value < 2.2e-16
## alternative hypothesis: true mean is less than 40
## 95 percent confidence interval:
##      -Inf 37.98979
## sample estimates:
## mean of x
##      37.76
```

P-value가 0에 가까우므로 귀무 가설을 기각하고 대립가설을 채택할 수 있다. 다시 말해, 응답자의 평균 나이가 40세 이하라는 주장은 통계적으로 유의하다.

Question 4

```
adresstypeTable = table(email.df$adresstype)
round(prop.table(adresstypeTable), 3)
```

```
##
## rural urban
## 0.373 0.627
```

Question 5

```
binom.test(373, 1000, 0.36, alternative = "less")
```

```
##
## Exact binomial test
##
## data: 373 and 1000
## number of successes = 373, number of trials = 1000, p-value = 0.8133
## alternative hypothesis: true probability of success is less than 0.36
## 95 percent confidence interval:
## 0.0000000 0.3989089
## sample estimates:
## probability of success
## 0.373
```

P-Value가 0.8133로 귀무가설을 기각할 수 없으므로, 지방(rural) 인구 비율이 36% 이하일 것이라는 대립가설은 통계적으로 유의하지 않다.

Question 6

```
respondByAddress = aggregate(
  respond ~ adresstype,
  data = email.df,
  function(x) { round(mean(x), 3) }
)
respondByAddress
```

```
## adresstype respond
## 1 rural 0.123
## 2 urban 0.144
```