# 2014년 주거실태를 이용한

# 분석보고서

## 20196384 유나현

목차

# 서울과 경기도 통근시간 차이 분석 (t test)

# 강남구 수성구 해운대구 주택 가격 차이 분석 (anova)

# 현재 주택 면적에 영향을 주는 변수 분석 (regression)

# 서울과 경기도 통근시간 차이 분석 (t test)R소스코드

## R소스코드

## install.packages("dplyr")

## install.packages("readxl")

## install.packages("ggplot2")

## library(ggplot2)

## library(readxl)

## library(dplyr)

## df<-read\_excel("2014년도 주거실태조사\_공표자료(시군구 명칭 수정).xlsx",sheet="조사결과")

## df2 <- df %>% select(sigungu,q20\_1\_a,q8\_5\_a,q50\_1,q49\_5,q53\_1\_4,dq2\_4\_tm,q12\_1,q52\_4,sido,dq2\_3,q12\_1,q37\_1\_a)

## df2 <- rename(df2,area\_now=q20\_1\_a,

## area\_first=q8\_5\_a,

## house\_money=q12\_1,

## commute\_time=dq2\_4\_tm,

## living\_expense=q50\_1,

## debt=q53\_1\_4,

## asset=q52\_4,

## income=q49\_5,

## trans=dq2\_3,

## house\_money=q12\_1,

## area\_want=q37\_1\_a)

## df2$sido<-ifelse(df2$sido==11,"서울특별시",ifelse(df2$sido==21,"부산광역시",ifelse(df2$sido==22,"대구광역시",ifelse(df2$sido==23,"인천광역시",ifelse(df2$sido==24,"광주광역시",ifelse(df2$sido==25,"대전광역시",ifelse(df2$sido==26,"울산광역시",ifelse(df2$sido==29,"세종특별자치시",ifelse(df2$sido==31,"경기도",ifelse(df2$sido==32,"강원도", ifelse(df2$sido==33,"충청북도",ifelse(df2$sido==34,"충청남도",ifelse(df2$sido==35,"전라북도",ifelse(df2$sido==36,"전라남도",ifelse(df2$sido==37,"경상북도",ifelse(df2$sido==38,"경상남도",ifelse(df2$sido==39,"제주도","기타")))))))))))))))))

## #1 서울 경기도 통근시간 차이

## commute<-df2 %>%

## select(sido,commute\_time) %>%

## filter(sido %in% c("경기도","서울특별시"))

## commute=na.omit(commute)

## var.test(commute\_time~sido,data=commute)

## t.test(commute\_time~sido,data=commute,var.equal=T)

## 자료분석 결과물과 해석

F test to compare two variances

data: commute\_time by sido

F = 0.9512, num df = 3092, denom df = 2708, p-value = 0.1785

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.8841901 1.0231113

sample estimates:

ratio of variances

0.9512027

> t.test(commute\_time~sido,data=commute,var.equal=F)

Welch Two Sample t-test

data: commute\_time by sido

t = -7.8162, df = 5659.4, p-value = 6.449e-15

alternative hypothesis: true difference in means between group 경기도 and group 서울특별시 is not equal to 0

95 percent confidence interval:

-6.085906 -3.645224

sample estimates:

mean in group 경기도 mean in group 서울특별시

31.87973 36.74529

서울과 경기도의 통근시간이 차이가 있는지 분석한 결과는 다음과 같다. 귀무가설은 서울의 통근시간과 경기도의 통근시간이 같다이고 대립가설은 같지 않다 이다.

분산의 동일성 검정을 시행한 결과 P-VALUE = 0.1785 으로 귀무가설(분산이 같다)를 기각할 수 없어 var.equal=F 로 두고 t검정을 시행한다.

분석결과 P-VALUE = 6.449E-15이므로 귀무가설을 기각하여 서울과 경기도의 통근시간이 차이가 있는 것으로 보인다. 경기도의 통근시간의 평균은31.87973(분) 이고 서울특별시의 통근시간의 평균은36.74529(분)이다.

# 강남구 수성구 해운대구 주택 가격 차이 분석 (anova)

## R소스코드

## #2 주택 가격 차이

## sido\_price<-df2 %>%

## select(sigungu,house\_money) %>%

## filter(sigungu %in% c('강남구', '수성구', '해운대구'))

## sido\_price = na.omit(sido\_price)

## bartlett.test(house\_money~sigungu,data=sido\_price)

## oneway.test(house\_money~sigungu, data=sido\_price,var.equal = T)

## 자료분석 결과물과 해석

Bartlett test of homogeneity of variances

data: house\_money by sigungu

Bartlett's K-squared = 268.43, df = 2, p-value < 2.2e-16

> oneway.test(house\_money~sigungu, data=sido\_price,var.equal = T)

One-way analysis of means

data: house\_money and sigungu

F = 104.02, num df = 2, denom df = 286, p-value < 2.2e-16

서울의 강남구, 대구의 수성구, 부산의 해운대구의 주택 가격을 분석한 결과는 다음과 같다.

귀무가설은 강남구, 수성구, 해운대구의 주택가격의 차이가 없다이고 대립가설은 차이가 있다 이다.

분산분석을 시행하기 전에 분산의 동일성 검정을 한 결과 P-VALUE < 2.2E-16이므로 귀무가설(분산이 같다)를 기각하여 분산이 동일하다고 볼 수 있으므로 var.equal=T 로 두고 분산분석을 시행한다.

분석결과 P-VALUE < 2.2E-16이므로 귀무가설을 기각하여 강남구, 수성구, 해운대구의 주택가격의 차이가 있다고 볼 수 있다.

# 현재 주택 면적에 영향을 주는 변수 분석 (regression)

## R소스코드

## #3 현재 주택 면적에 영향을 주는 변수

## df2$debt <- ifelse(df2$debt == 9999999, NA, df2$debt)

## df2$income <- ifelse(df2$income == 9999999, NA, df2$income)

## df2$asset <- ifelse(df2$asset == 9999999, NA, df2$asset)

## df2$house\_money <- ifelse(df2$house\_money == 9999999, NA, df2$house\_money)

## df2$area\_now <- ifelse(df2$area\_now == 999, NA, df2$area\_now)

## df2$area\_first <- ifelse(df2$area\_first == 999, NA, df2$area\_first)

## df2$area\_want <- ifelse(df2$area\_want == 999, NA, df2$area\_want)

## df2$living\_expense <- ifelse(df2$living\_expense == 9999, NA, df2$living\_expense)

## money<-df2 %>%

## select(area\_first,area\_now,asset,house\_money,income,debt,area\_want,living\_expense) %>%

## filter(!is.na(asset)) %>%

## filter(!is.na(income)) %>%

## filter(!is.na(house\_money)) %>%

## filter(!is.na(debt)) %>%

## filter(!is.na(area\_first)) %>%

## filter(!is.na(area\_now)) %>%

## filter(!is.na(living\_expense))

## plot(area\_now~area\_first,data=money)

## plot(area\_now~debt,data=money)

## plot(area\_now~income,data=money)

## plot(area\_now~asset,data=money)

## plot(area\_now~house\_money,data=money)

## plot(area\_now~living\_expense,data=money)

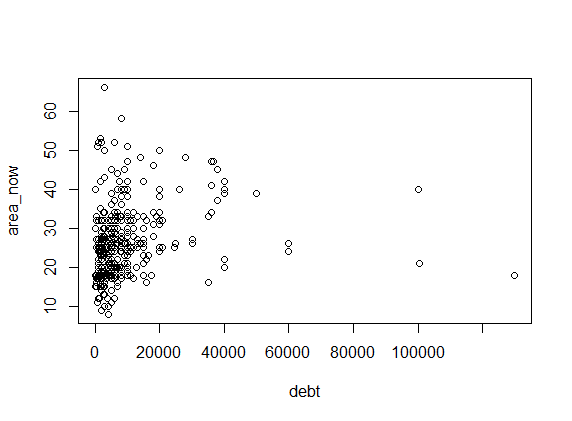
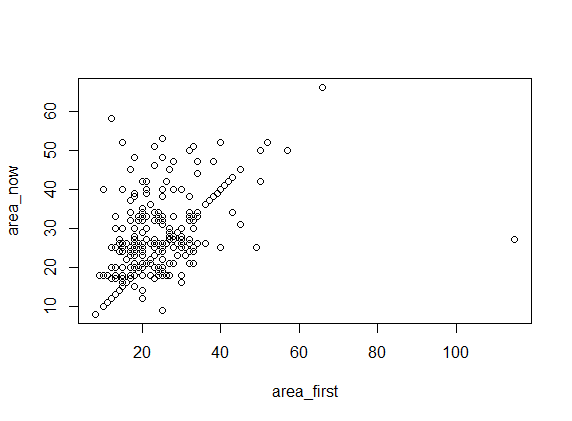
## regression1=lm(area\_now~area\_first+income+asset+house\_money+debt+living\_expense,data=money)

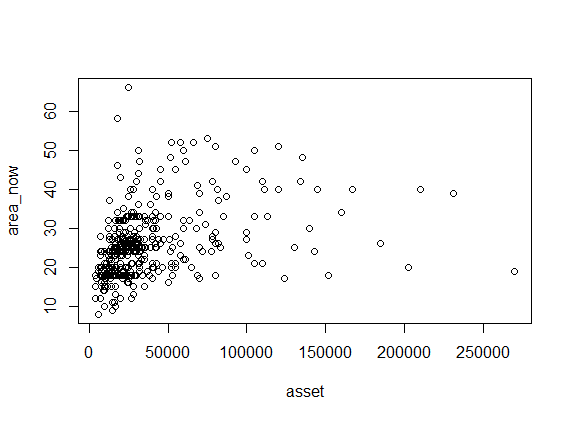
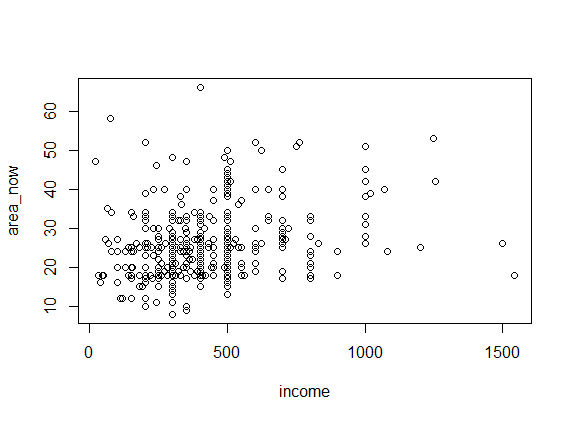
## summary(regression1)

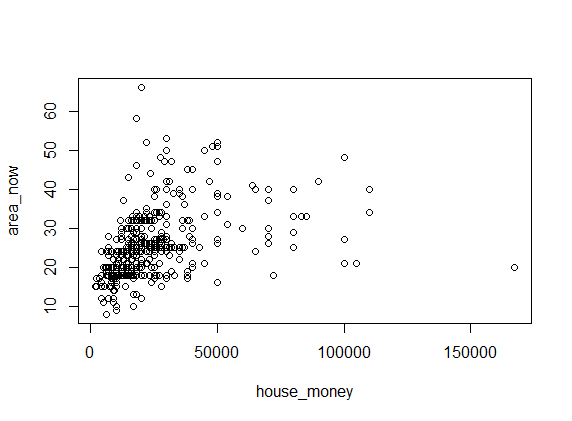
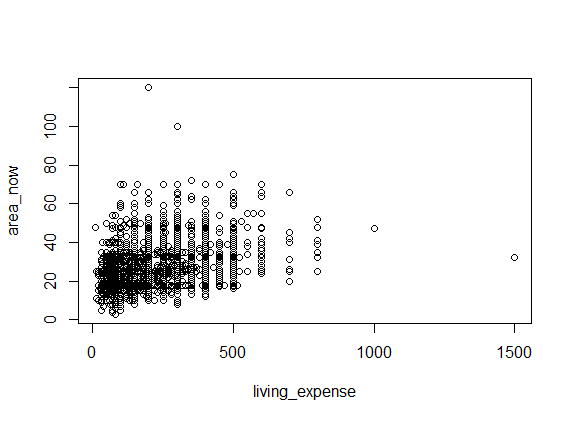
## regression2=lm(area\_now~area\_first+asset+house\_money+debt+living\_expense,data=money)

## summary(regression2)

## 자료분석 결과물과 해석





Residuals:

Min 1Q Median 3Q Max

-63.778 -3.609 -0.852 2.556 48.686

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 8.867e+00 3.394e-01 26.127 < 2e-16 \*\*\*

area\_first 4.816e-01 1.204e-02 40.009 < 2e-16 \*\*\*

income 4.850e-04 6.771e-04 0.716 0.47390

asset 1.181e-05 3.794e-06 3.112 0.00187 \*\*

house\_money 1.144e-04 7.716e-06 14.824 < 2e-16 \*\*\*

debt 3.343e-05 1.128e-05 2.963 0.00306 \*\*

living\_expense 1.103e-02 1.229e-03 8.977 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.903 on 4580 degrees of freedom

Multiple R-squared: 0.4171, Adjusted R-squared: 0.4163

F-statistic: 546.1 on 6 and 4580 DF, p-value: < 2.2e-16

Residuals:

Min 1Q Median 3Q Max

-63.853 -3.629 -0.851 2.560 48.648

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 8.877e+00 3.391e-01 26.179 < 2e-16 \*\*\*

area\_first 4.818e-01 1.203e-02 40.045 < 2e-16 \*\*\*

asset 1.222e-05 3.750e-06 3.257 0.00113 \*\*

house\_money 1.146e-04 7.708e-06 14.873 < 2e-16 \*\*\*

debt 3.415e-05 1.124e-05 3.039 0.00239 \*\*

living\_expense 1.161e-02 9.261e-04 12.536 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.903 on 4581 degrees of freedom

Multiple R-squared: 0.417, Adjusted R-squared: 0.4164

F-statistic: 655.3 on 5 and 4581 DF, p-value: < 2.2e-1

현재 주택 면적에 영향을 주는 변수들에 대한 회귀분석을 시행하였다. 우선 월 평균 가구소득(income), 가구의 총 자산(asset), 가구의 총 부채(Debt), 생애 최초 마련한 주택의 면적(area\_first), 현재 살고 있는 주택의 가격(house\_money), 월 평균 총 생활비(living\_expense)를 변수로 가져와 각각 결측치를 제거한 후 현재 주택 면적과의 그래프를 그린결과 어느정도 상관관계가 보여 회귀분석을 실시하였다.

귀무가설은 회귀식이 의미가 없다이고 대립가설은 회귀식이 의미가 있다이다.

regression1에서 모든 변수들을 넣어 회귀분석을 시행한 결과 income의 p-value가 유의하지 않아 regression2에서는 income변수를 제외하고 다시 회귀분석을 시행하였다. 그 결과 모든 변수가 유의하였고 P-VALUE: < 2.2E-1으로 귀무가설을 기각하여 회귀모형이 유의하다는 것을 알 수 있다. 회귀식은 다음과 같이 나타낼 수 있고 y=8.877e+00+4.818e-01AF+1.222e-05A+1.146e-04HM+3.415e-05D+1.161e-02LE, 결정계수는 0.417로 해당 회귀식으로 전체 변동의 41.7%가 설명된다.