

# 2014 년 주거실태를 이용한 분석보고서

20196384 유나현

## 1. 서울과 경기도 통근시간 차이 분석 (T TEST)

R 소스코드

```
install.packages("dplyr")
```

```
install.packages("readxl")
```

```
install.packages("ggplot2")
```

```
library(ggplot2)
```

```
library(readxl)
```

```
library(dplyr)
```

```
df<-read_excel("2014 년도 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 같다고 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 T 검정을 시행한다.

분석결과 P-VALUE = 6.449E-15 이므로 귀무가설을 기각하여 서울과 경기도의 통근시간이 차이가 있는 것으로 보인다. 경기도의 통근시간의 평균은 31.87973(분) 이고 서울특별시의 통근시간의 평균은 36.74529(분)이다.

## 2. 강남구 수성구 해운대구 주택 가격 차이 분석 (ANOVA)

### R 소스코드

## #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 있다이다.

분산분석을 시행하기 전에 분산의 동일성 검정을 한 결과 P-VALUE < 2.2E-16 이므로

귀무가설(분산이 같다)를 기각하여 분산이 동일하다고 볼 수 있으므로 VAR.EQUAL=T 로 두고 분산분석을 시행한다.

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

### 3. 현재 주택 면적에 영향을 주는 변수 분석 (REGRESSION)

#### R 소스코드

```
#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)
```

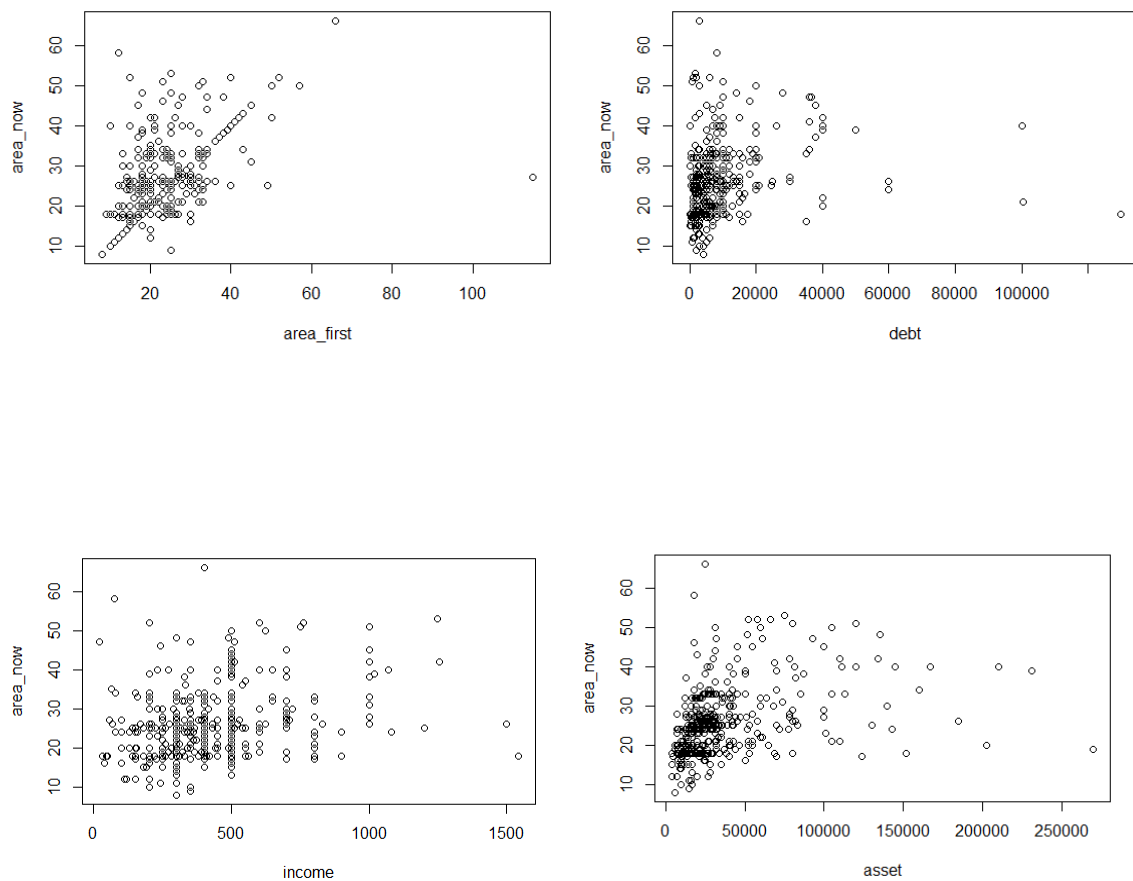
```
regression1=lm(area_now~area_first+income+asset+house_money+debt+living_expense,data  
=money)
```

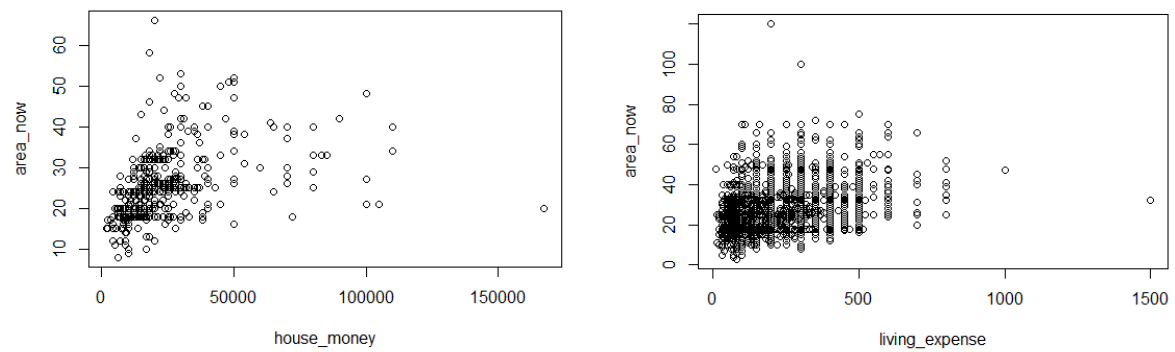
```
summary(regression1)
```

```
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.

REGRESSION1 에서 모든 변수들을 넣어 회귀분석을 시행한 결과 INCOME 의 P-VALUE 가 유의하지 않아 REGRESSION2 에서는 INCOME 변수를 제외하고 다시 회귀분석을 시행하였다. 그 결과 모든 변수가 유의하였고 P-VALUE: < 2.2E-1 으로 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 0.417 로 해당 회귀식으로 전체 변동의 41.7%가 설명된다.