Go To College (Dataset)
Group 10 - Section 6




NAME	MUHAMMAD NAJWAN HAZIM BIN KHAIRI	MUHAMMAD TAUFIQ BIN JURIMI	AFIQ FAHMI BIN ROSLAN	AZHAN HANIFF BIN AZNI
STUDENT ID	A21EC0087	A21EC0095	A21EC0153	A21EC0017

HYPOTHESIS TESTING

```
• data <- read_csv("data.csv")
• View(data)
• x=data$average_grades
• mean(x)
• sd(x)
• n=1000
• sd=3.378738
• xbar=86.0972
• mu=84
• alpha=0.05
• z=(xbar-mu)/(sd/sqrt(n))
• z.alpha=qnorm(1-alpha)
```

```
z= 19.6284195723525
z.alpha = 1.64485362695147
```

H_0 = μ= 84
H_1 = μ>84

- $\bar{X}$ = 86.0972
- μ = 84
- S = 3.378738
- N = 1000

test statistics, Z_0 =19.6284195723525
Z_a = 1.64485362695147

H_0 is rejected

Since Z_0 =19.628 >Z = 1.6448, we reject H_0, there is sufficient evidence that the mean of grade exceeds 84.

CORRELATION ANALYSIS

```
> cor(house_area,average_grades)
[1] 0.4095654
```

```
> n <- 1000
> r <- cor(house_area,average_grades)
> t <- r/(sqrt((1-(r^2))/(n-2)))
```

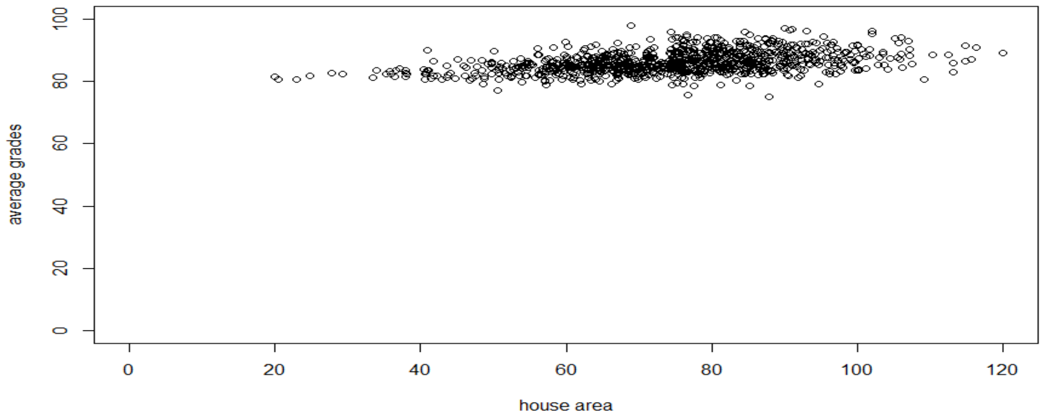
n	1000
r	0.409565396175046
t	14.1827357407675

```
> cor.test(house_area,average_grades,method="pearson")

        Pearson's product-moment correlation

data:  house_area and average_grades
t = 14.183, df = 998, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3566272 0.4598820
sample estimates:
      cor
0.4095654
```

```
> plot(house_area,average_grades, xlim=c(0,120), ylim=c(0,100), xlab="house area",
ylab="average grades")
```



$r = 0.4095654$

Hypothesis Statement :

$H_0 = \rho = 0$ (no linear correlation)

$H_1 = \rho \neq 0$ (linear correlation exists)

$t = 14.1827$

Lower tail critical value $-t(\alpha/2=0.025, df=998) = -1.9623$

Upper tail critical value $t(\alpha/2=0.025, df=998)= 1.9623$

Since test statistics $t = 14.1827 > 1.9623$

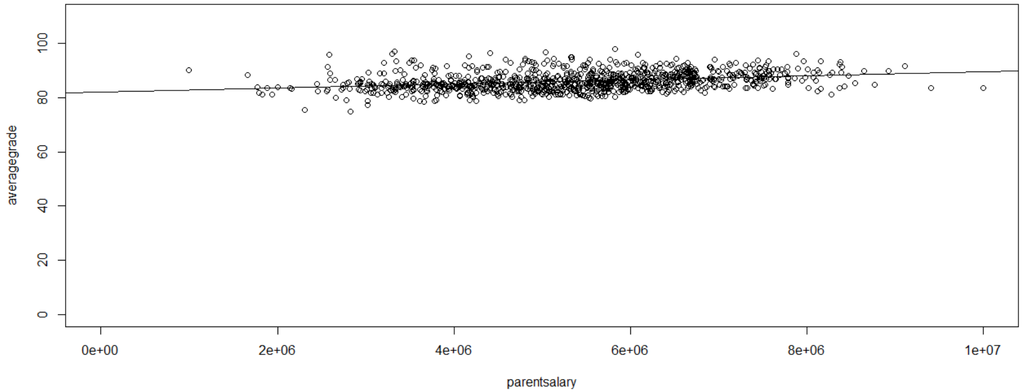
H_0 is rejected

There is sufficient evidence that supports H_1 that claims that there exists a linear correlation between **house area**(**house_area**) and **average grades**(**average_grades**).

REGRESSION ANALYSIS

```
• parentsalary <- data[,7]
• > averagegrade <- data[,9]
• > cor(parentsalary, averagegrade)
• > model<-
  lm(averagegrade~parentsalary)
• > plot(parentsalary,
  averagegrade, xlim=c(0,10000000),
  ylim=c(0,110))
• > abline(model)
• > summary(model)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.211e+01  4.050e-01  202.74   <2e-16 ***
parentsalary 7.415e-07  7.284e-08   10.18   <2e-16 ***
```



$\hat{y} = 82.11 + 0.0000007415x$

Intersection Coefficient, $b_0 = 82.11$

Interpretation of the Slope Coefficient, $b_1 = 0.0000007415$

$R = 0.3067119$

$R^2 = 0.09407$

$H_0 : \beta 1 = 0$

$H_1 : \beta 1 \neq 0$

Critical value,

$-t_{0.025, 998} = -1.962$

$t_{0.025, 998} = 1.962$

$s_{b1} = 0.00000007284$

$t = 10.18$

Since $t_0 = 10.18 > t = 1.962$, we reject H_0

There is sufficient evidence that parent salary affects students' average grade.

CHI SQUARE TEST OF INDEPENDENCE

H0 : the type of school is independent to the residence.
H1 : the type of school is dependent to the residence.

$x^2 = 60.336$

Critical value = 3.84145882069413

P(Critical Value) = 0.000000000000007997

Test statistics fall within the critical value, thus we reject the null hypothesis

	Residence		
Type of school	Rural	Urban	**Total**
Academic	221	388	609
Vocational	240	151	391
Total	461	539	1000

```
> View(data)
> tbl = table(data$type_school, data$residence)
>
> chisq.test(tbl, correct=FALSE)

        Pearson's Chi-squared test

data:  tbl
X-squared = 60.336, df = 1, p-value = 7.997e-15

> alpha <- 0.05
> x2.alpha <- qchisq(alpha, df=1, lower.tail = FALSE)
> tbl

              Rural Urban
  Academic      221   388
  Vocational    240   151
> |
```