Gabriella Weis
Mr. Lewis
H Applied Statistics
18 June 2024

# #9: First Generation College Attendee

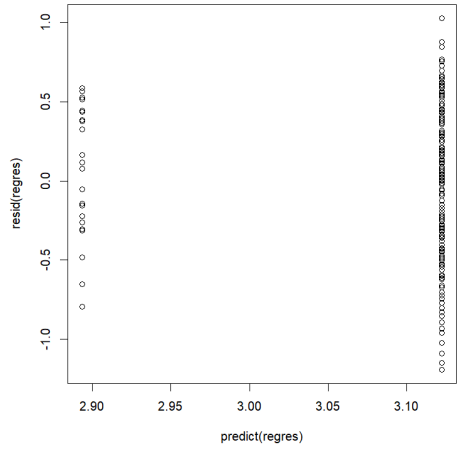
a) 2-Sample T-test:

$H_o$: $\mu_1 - \mu_2 = 0$
$H_a$: $\mu_1 - \mu_2 \neq 0$
$\alpha = 0.05$

$\mu_1$= Average GPA of students who are first gen. (1)
$\mu_2$= Average GPA of students who are not first gen. (0)

Conditions:

- Random: Assume random selection of students.
- Normality:
  - $\mu_1$: Probability plot is not linear. Proceed with caution.




  - $\mu_2$: CLT (N > 30, 194 > 30)
- Independence: Assume independence.

Calculations:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{(2.893600 - 3.122268) - (0)}{\sqrt{\frac{0.4365001^2}{25} + \frac{0.4637245^2}{194}}} = 2.4474$$

p-value = p(t > 2.4474)x2 = 0.02017

Since the p-value (0.02017) < α (0.05), we reject the null hypothesis. Therefore, there is convincing evidence that the difference between the average GPA of students who are first generation and the average GPA of students who are first generation is not equal to zero.

(Direction of the relationship is negative.)

Call output:

```
data:  GPA by FirstGen
t = 2.4474, df = 31.406, p-value = 0.02017
alternative hypothesis: true difference in means between group 0 and group 1 is
not equal to 0
95 percent confidence interval:
 0.03820979 0.41912630
sample estimates:
mean in group 0 mean in group 1
       3.122268        2.893600
```

b). Simple linear regression: predict GPA using FirstGen.

Model Utility Test:

$H_o$: β = 0

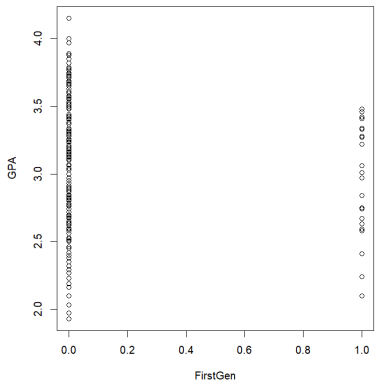
$H_a$: β ≠ 0

α = 0.05

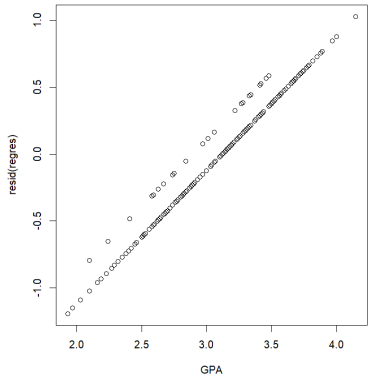
β = change in GPA with every 1 unit increase in FirstGen.

Conditions:

- Linear: Not linear (scatter plot). Proceed with caution.




- Independence: Assume Independence.
- Normality: CLT (N > 30, 219 > 30)
- Equal Variance: Residuals are not randomly distributed (residual plot). Proceed with caution.




Calculations:

$t = \frac{b-\beta}{S_b}$ w/df = n-2

$t = \frac{-0.22867-0}{0.09792}$ = -2.3353 w/df = 217

p-value = p(t > -2.3353)x2 = 0.02044

Since the p-value (0.02044) < α (0.05), we reject the null hypothesis. Therefore, there is convincing evidence that the true change in GPA with every 1 unit increase in FirstGen. is not equal to zero.

There is a weak, negative, nonlinear relationship between FirstGen and GPA.

(Direction of the relationship is negative.)

c). Logistic Regression: predict FirstGen. using GPA

x = GPA
β = change in FirstGen. with every 1 unit increase in GPA

Odds of FirstGen. based off on GPA
$ln(\frac{p}{1-p}) = 1.0751 - 1.0381x$

Probability of FirstGen. based on GPA
$p = \frac{e^{1.0751-1.0381x}}{e^{1.0751-1.0381x}+1}$

Model Utility Test hypotheses and results:

$H_o$: β = 0
$H_a$: β ≠ 0
α = 0.05

p-value = 0.023

Since the p-value (0.023) < α (0.05), we reject the null hypothesis. Therefore, there is convincing evidence that the true change in FirstGen. with every 1 unit increase in GPA is not equal to zero.

Both the t-test in part a and the Model Utility Test in part b also reject the null hypothesis with p-values less than 0.05.

d). Questions:

What is the probability of being a first generational college attendee given a 4.0 GPA?

$p = \frac{e^{1.0751-1.0381(4)}}{e^{1.0751-1.0381(4)}+1} = 0.04405 = 4.4\%$

What GPA will be needed to assure a predicted p(first generational college attendee) = 0.5?

$ln(\frac{0.5}{1-0.5}) = 1.0751 - 1.0381x$
$0 = 1.0751 - 1.0381x$

$-\ 1.0751\ =\ \ -\ 1.0381x$
x = 1.0356

A 1.0356 GPA.

If a student is selected randomly, and has a predicted 2:1 odds of being a first generational college attendee, what is their GPA?

$\frac{2}{1} = e^{1.0751-1.0381x}$
$ln(2)\ =\ 1.0751\ -\ 1.0381x$
$ln(2)\ -\ 1.0751\ =\ \ -\ 1.0381x$
x = 0.3679

Their GPA is 0.3679.

e). Best multiple regression model to predict first year college GPA: ($\widehat{y}$ = predicted GPA)

| Model | $R^2$(adj.) | Se | F-statistic | p-value |
|---|---|---|---|---|
| $\widehat{y}$ = 3.12227 - 0.22867(FirstGen) | 0.02002 | 0.4608 | 5.454 | 0.02044 |
| $\widehat{y}$ = 2.84239 + 0.32125(White) | 0.07515 | 0.4476 | 18.71 | 0.00002316 |
| $\widehat{y}$ = 3.07325 + 0.04920(Male) | -0.001802 | 0.4659 | 0.6078 | 0.4365 |
| $\widehat{y}$ = 1.15966 - 0.27168(FirstGen) + 0.56984(HSGPA) | 0.2271 | 0.4092 | 33.02 | $\approx$ 0 |
| $\widehat{y}$ = 0.8449120 - 0.5499982(HSGPA) + 0.0005995(SATM) - 0.2450794(First | 0.2324 | 0.4078 | 23.01 | $\approx$ 0 |

| | | | | |
|---|---|---|---|---|
| Gen) | | | | |
| $\hat{y}$ = 1.00377 + 0.55223(HSGPA) - 0.19034(FirstGen) + 0.26256(White) | 0.274 | 0.3966 | 28.43 | $\approx$ 0 |

Chosen Model:

$\hat{y}$ = 1.00377 + 0.55223(HSGPA) - 0.19034(FirstGen) + 0.26256(White)