This Portfolio contains 6 research questions focusing on the predictors of graduation rates of colleges. The result is that the Students-Faculty Ratio, the percentage of new students from the top 10% of high school class, Acceptance Rate, and Ratio of Faculty with Ph.D.'s affect Graduation Rate in various ways. I applied 6 models: Pearson Correlation, Dependent Correlation, Simple Linear Regression (SLR), Multiple Linear Regression (MLR), Moderation and Mediation. The dataset consists of 777 colleges and its 20 features.

Research Question 1:

The research question is whether the correlation between graduation rate and students and faculty ratio were negatively correlated.

1. Research Description:

The dataset contains a random sample of 777 universities and their values for graduate rate and students/faculty ratio. The alpha level for this research question is 0.05.

2. Analysis:

(1) Null hypothesis:

H1: There is a negative relationship between the graduation rate and students and faculty ratio. Or H1: $\rho < 0$

H0: There is not a negative relationship between the graduation rate and students and faculty ratio. Or H0: $\rho <= 0$

(2) Critical value:

Df = 775, α = 0.05, one-tailed, critical t = -1.647

```
> qt(0.05, 775)
[1] -1.646822
```

(3) Sample test analysis:

r = -0.307, t(775)= -8.971

```
> cor.test(Grad.Rate, S.F.Ratio, method = "pearson")

        Pearson's product-moment correlation

data:  Grad.Rate and S.F.Ratio
t = -8.9708, df = 775, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.3690817 -0.2415888
sample estimates:
       cor
-0.3067104
```

(4) Conclusion:

Reject the Null. There is a significant negative relationship between the students and faculty ratio and graduation rate. [$r = -0.307$, $t(775) = -8.971$, $p < 0.05$]

Research Question 2:

The research question is whether the correlation between student-faculty ratio and graduation rate was stronger for the private university than for the public.

1. Research Description:

The researcher obtained a random sample of values for 565 private universities and 212 public universities on student-faculty ratio and graduation rate. He tested his hypothesis using an alpha of 0.05.

2. Analysis:

(1) Null hypothesis:

H1: The correlation between student-faculty ratio and graduation rate was stronger for the private university than for the public.

Or H1: ρPri - ρPub > 0

H0: The correlation between student-faculty ratio and graduation rate was not stronger for the private university than for the public.

Or H0: ρPri - ρPub <= 0

(2) Critical value:

α = 0.05, one-tailed, critical z = 1.645

(3) Sample test analysis:

I used R software to calculate each group's correlation estimate

```
College%>% group_by(Private) %>% summarize(cor(S.F.Ratio,Grad.Rate))
# Public  1       0                    -0.0820
# Private 2       1                    -0.209
```

And then the r.test function in the psych library to calculate the sample's te st statistic:

```
> library(psych)
> r.test(n = n.Pri, n2 = n.Pub, -0.209, -0.082, twotailed = FALSE)
Correlation tests
Call:r.test(n = n.Pri, r12 = -0.209, r34 = -0.082, n2 = n.Pub, twotail
ed = FALSE)
Test of difference between two independent correlations
 z value -1.6    with probability  0.05
```

(4) Conclusion:

Fail to reject the null.

The correlation between student-faculty ratio and graduation rate was not stronger for the private university than for the public. [r-public = -0.082, n-public = 212, r-private = -0.209, n-private = 565, Z = - 1.6, p > 0.05]

Research Question 3:

The research question was whether the student-faculty ratio was a predictor of the graduation rate of a college.

1. Research Description:

The researcher hypothesized that the prediction relationship could be negative. Using the sample of data for the 777 universities across the US and α of 0.05.

2. Analysis:

(1) Null hypothesis:

H1: The student-faculty ratio is a negative predictor of the graduation rate of this college.

H0: The student-faculty ratio is not a negative predictor of the graduation rate of this college.

(2) Critical value:

α = 0.05, one-tailed, df = n – 2 = 775, t(df = 775) = -1.647

```
> qt(0.05, 775)
[1] -1.646822
```

(3) Sample test analysis:

```
> summary(lm(formula = Grad.Rate ~ S.F.Ratio))
Call:
lm(formula = Grad.Rate ~ S.F.Ratio)

Residuals:
    Min      1Q  Median      3Q     Max 
-54.443 -11.094   0.284  11.612  52.817 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  84.2168     2.1713  38.786   <2e-16 ***
S.F.Ratio    -1.3310     0.1484  -8.971   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16.36 on 775 degrees of freedom
Multiple R-squared:  0.09407,	Adjusted R-squared:  0.0929 
F-statistic: 80.48 on 1 and 775 DF,  p-value: < 2.2e-16
---
```

So the sample test statistics t=-8.971

To obtain the standardized coefficient estimates, we run the following in R

```
> lm(formula = scale(Grad.Rate) ~ scale(S.F.Ratio))
```

```
Call:
lm(formula = scale(Grad.Rate) ~ scale(S.F.Ratio))

Coefficients:
     (Intercept)  scale(S.F.Ratio)
       2.377e-16        -3.067e-01
```

(4) Conclusion:

Reject the null.

The student-faculty ratio is a negative predictor of the graduation rate of this college. [B = -1.331, β = -0.307, t(775)= -8.971, $p < 0.05$]

The result indicates that for two universities which differ by 1 point in student-faculty ratio is predicted to have 1.331 percentage decrease in the graduation rate. Also, the university with I standard deviation higher in 1 point in student-faculty ratio is predicted to have 0.307 standard deviation decrease in the graduation rate.

Research Question 4:

The researcher was interested in whether the percentage of new students from top 10% of high school class (Top10perc) and the student and faculty ratio (S.F.Ratio) were significant predictors of the graduation rate of a college (Grad.Rate).

1. Research Description:

Both predictors were mean centered. The researcher used the sample of data for 777 colleges and tested the hypothesis using α of 0.05.

2. Analysis:

(1) Null hypothesis:

H1A: The predictors (the percentage of new students from top 10% of high school class (Top10perc) and the student and faculty ratio (S.F.Ratio)) explain variability in the graduation rate of a college.

H0A: The predictors (the percentage of new students from top 10% of high school class (Top10perc) and the student and faculty ratio (S.F.Ratio)) do not explain variability in the graduation rate of a college.

H1B: The student and faculty ratio (while controlling for the percentage of new students from top 10% of high school class (Top10perc)) is a negative predictor of the graduation rate of a college.

H0B: The student and faculty ratio (while controlling for the percentage of new students from top 10% of high school class (Top10perc)) is not a negative predictor of the graduation rate of a college.

H1C: The percentage of new students from top 10 (while controlling for the student and faculty ratio) is a positive predictor of the graduation rate of a college.

H0C: The percentage of new students from top 10 (while controlling for the student and faculty ratio) is not a positive predictor of the graduation rate of a college.

H1D: The graduation rate of a college with the mean percentage of new students from top 10 and mean student and faculty ratio is greater than zero.

H0D: The graduation rate of a college with the mean percentage of new students from top 10 and mean student and faculty ratio is not greater than zero.

(2) Critical value:

```
# F critical for H0A = 3.007357
qf(0.95, 2, 774)
# t critical for H0B = -1.646825
qt(0.05, 774)
# t critical for H0C = 1.646825
qt(0.05, 774)
```

(3) Sample test analysis:

3-1. mean-center each of the two predictors:

```
> Top10perc.ctrd <-  scale(Top10perc, center = TRUE,scale = FALSE)
> S.F.Ratio.ctrd <-  scale(S.F.Ratio, center = TRUE, scale = FALSE)
```

Then, use:

```
> summary(lm(Grad.Rate ~ Top10perc + S.F.Ratio))

Call:
lm(formula = Grad.Rate ~ Top10perc + S.F.Ratio)

Residuals:
    Min      1Q  Median      3Q     Max 
-47.619  -9.990   0.121   9.343  60.658 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  61.9293     2.5876  23.933  < 2e-16 ***
Top10perc     0.4309     0.0326  13.216  < 2e-16 ***
S.F.Ratio    -0.5920     0.1453  -4.074  5.1e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.79 on 774 degrees of freedom
Multiple R-squared:  0.2609,	Adjusted R-squared:  0.259 
F-statistic: 136.6 on 2 and 774 DF,  p-value: < 2.2e-16
```

And use:

```
# 3-3. mean centered and standardized multiple linear regression model:

lm(scale(Grad.Rate) ~ scale(Top10perc.ctrd) + scale(S.F.Ratio.ctrd))

# Coefficients:
# (Intercept)  scale(Top10perc.ctrd)  scale(S.F.Ratio.ctrd)
# 1.978e-16            4.425e-01             -1.364e-01
```

(4) Conclusion:

For testing H0A:

Reject the null.

The predictors (the percentage of new students from top 10% of high school class (Top10perc) and the student and faculty ratio (S.F.Ratio)) explain variability in the graduation rate of a college.

[R-squared = 0.261, R-squared adj = 0.259, F(2, 774) = 136.6. P <0.05].

Together, the percentage of new students from top 10 and the student and faculty ratio explain about 26.1% of the variability in the graduation rate of the college.

For testing H0B:
Reject the null.
The student and faculty ratio (while controlling for the percentage of new students from top 10) is a significant negative predictor of the graduation rate of a college.

[B = -0.592, β = -0.136, t(774) = -4.074, p < .05]

The results indicate that, controlling for the percentage of new students from top 10%, the larger student and faculty ratio is, the less the graduation rate of a college will be. Specifically, for two colleges with equal percentage of students from top 10%, for the college with one point higher in student and faculty ratio, that college will have 0.592 points lower in the graduation rate.

For testing H0C:

Reject the null.
The percentage of new students from top 10 (while controlling for the student and faculty ratio) is a significant positive predictor of the graduation rate of a college.

[B = 0.431, β = -0.443, t(774) = 13.216, p < .05]

The results indicate that, controlling for the student and faculty ratio, the more percentage of new students from top 10%, the more the graduation rate of a college will be. Specifically, for two colleges with equal student and faculty ratio, for the college with one percent higher in students from top 10%, that college will have 0.431 points higher in the graduation rate.

H0D
Reject the null.
The graduation rate of a college with the mean percentage of new students from top 10% and mean student and faculty ratio is significantly greater than zero.

[B = 61.929, t(45) = 23.933, p < .05]

The intercept is interpreted as the predicted graduation rate of a college (estimated to be 61.929) for some college at the mean on new students from top 10 and mean on the student and faculty ratio scale.

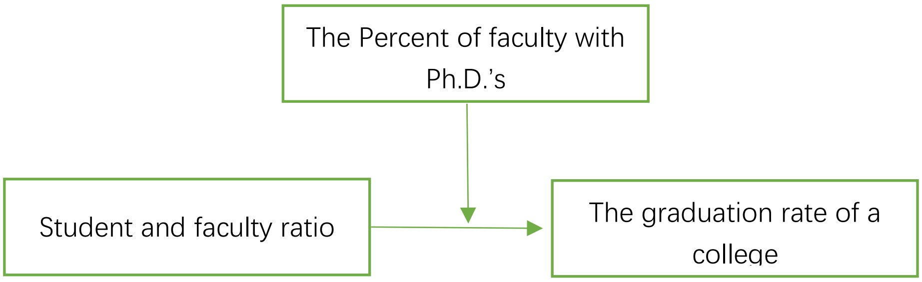
Research Question 5:

The research question was whether the relationship between the student and faculty ratio (S.F.Ratio) and the graduation rate of a college (Grad.Rate) was moderated by Percent of faculty with Ph.D.'s (FPhD).

1. Research Description:

The research test the hypothesis using a random sample of 777 colleges and α of 0.05.

The researcher was interested in testing the following model:



2. Analysis:

(1) Null hypothesis:

H1A: The ratio of faculty with PhD moderates the relationship between student and faculty ratio and the graduation rate of a college.

H0A: The ratio of faculty with PhD does not moderate the relationship between student and faculty ratio and the graduation rate of a college.

(2) Critical value:

α = 0.05, two-tailed, df =777-3-1= 773,

Critical t (773) = 1.963

```
> qt(0.975, (773))
[1] 1.963038
```

(3) Sample test analysis:

First, mean-center each of the predictor and moderator:

```
FPhD.ctrd <-  scale(FPhD, center = TRUE,scale = FALSE)
S.F.Ratio.ctrd <-  scale(S.F.Ratio, center = TRUE, scale = FALSE)
```

Build the model, test the Hypothesis and obtain the unstandardized slope.

```
summary(lm(Grad.Rate ~ S.F.Ratio.ctrd+FPhD.ctrd + S.F.Ratio.ctrd :FPhD.ctrd))

Residuals:
    Min      1Q  Median      3Q     Max 
-53.939  -9.680   0.777  10.712  65.953 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)              65.263487   0.564145 115.686  < 2e-16 ***
S.F.Ratio.ctrd           -1.201634   0.143110  -8.397  < 2e-16 ***
FPhD.ctrd                 0.260893   0.035402   7.369 4.41e-13 ***
S.F.Ratio.ctrd:FPhD.ctrd -0.023717   0.007592  -3.124  0.00185 **

Residual standard error: 15.62 on 773 degrees of freedom
Multiple R-squared:  0.1759,	Adjusted R-squared:  0.1727
F-statistic:    55 on 3 and 773 DF,  p-value: < 2.2e-16
```

4. Obtain the standardized slope.

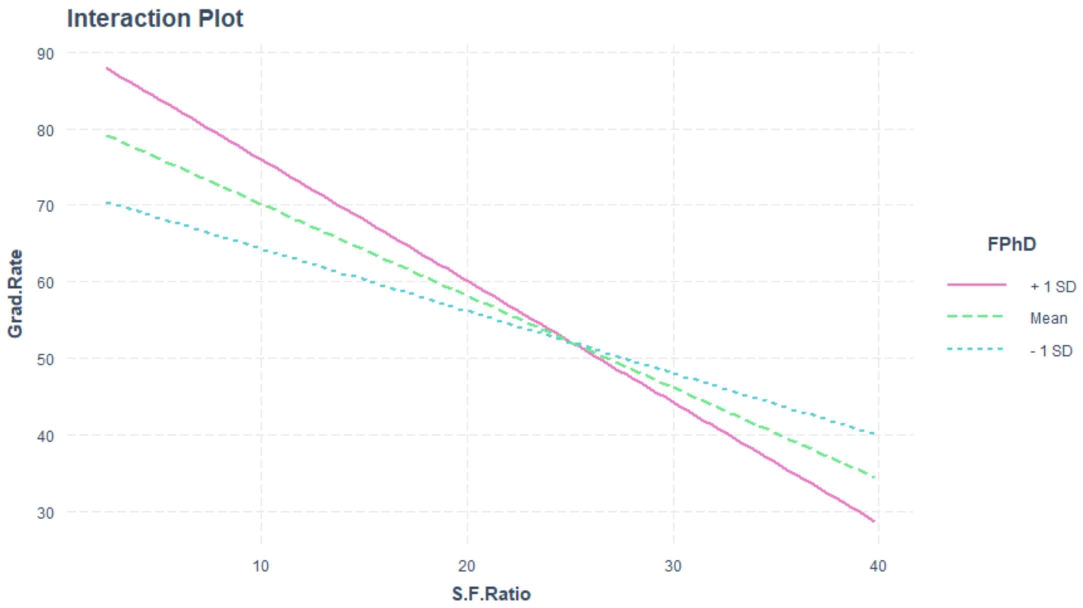
```
lm(scale(Grad.Rate) ~ scale(S.F.Ratio.ctrd) + scale(FPhD.ctrd) + scale(S.F.Ratio.ctrd * FPhD.ctrd),data = College)
Call:
lm(formula = scale(Grad.Rate) ~ scale(S.F.Ratio.ctrd) + scale(FPhD.ctrd) +
    scale(S.F.Ratio.ctrd * FPhD.ctrd), data = College)

Coefficients:
 (Intercept) scale(S.F.Ratio.ctrd) scale(FPhD.ctrd) scale(S.F.Ratio.ctrd * FPhD.ctrd)
 2.202e-16       -2.769e-01            2.480e-01                 -1.043e-01
```

(4) Conclusion:

The result led me to reject H0 and infer that the ratio of faculty with PhD's significantly moderates the relationship between student and faculty ratio and the graduation rate of a college.

[B =-0.024, β= -0.104, t(176) = -3.124, p < .05]

```
> # 5. Visualize the moderation
> install.packages("interactions")
> library(interactions)
> Grad.Rate.out <- lm(Grad.Rate ~ S.F.Ratio.ctrd+FPhD.ctrd + S.F.Ratio.ctrd * FPhD.ctrd,data = College)
> interact_plot(Grad.Rate.out, pred = S.F.Ratio.ctrd, modx = FPhD.ctrd, colors = c("#42c5f4","#54f284","#f45dcc"), main.title = "Interaction Plot")
```



The graph depicts predicted Graduation rate scores given student and faculty ratio for those with high, medium and low ratio of faculty with a PhD degree (where high, medium and low represented those with scores on the ratio of faculty with a PhD degree that were one standard deviation above the mean, at the mean and one standard deviation below the mean, respectively).

The pattern of the results indicates that the prediction of graduation rate by student and faculty ratio depends on the ration of faculty with a PhD degree. More specifically, the negative relationship between graduation rate and sf becomes less strong the fewer faculty with a PhD degree in a college. When the ratio of the faculty with a PhD degree is getting higher, the negative relationship between the graduation rate and student-faculty ratio will get stronger.

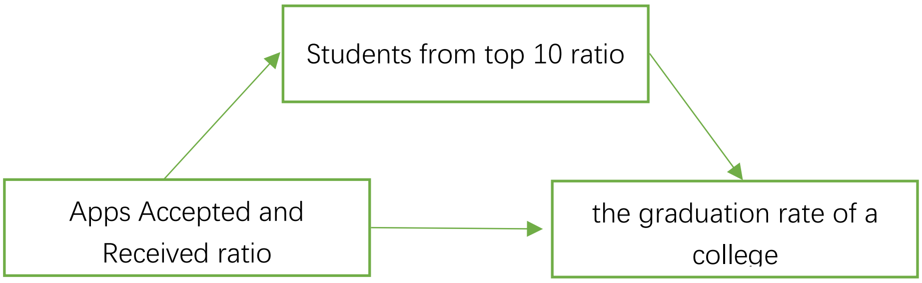
Research Question 6:

The researcher was interested in investigating whether The ratio of the applicants accepted and received by a college (AccRate) affected the graduation rate of a college (Grad.Rate) through the indirect effect on percentage of new students from top 10% of high school class (Top10perc).

1. Research Description:

The research test the hypothesis using a random sample of 777 colleges and α of 0.05.

The researcher was interested in testing the following model:



2. Analysis:

(1) Null hypothesis:

H1A: The ratio of the applicants accepted and received by a college affects a college's graduation rate through the indirect effect on the ratio of top 10 students enrolled in this college.

H0A: The ratio of the applicants accepted and received by a college does not affect a college's graduation rate through the indirect effect on the ratio of top 10 students enrolled in this college.

(2) Critical value:

α = 0.05, two-tailed

(3) Sample test analysis:

First, I obtained the a coefficient

```
> summary(lm(Top10perc ~ AccRate))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   70.433      2.879   24.46   <2e-16 ***
AccRate      -57.402      3.782  -15.18   <2e-16 ***

Residual standard error: 15.5 on 775 degrees of freedom
Multiple R-squared:  0.2291,	Adjusted R-squared:  0.2281
F-statistic: 230.4 on 1 and 775 DF,  p-value: < 2.2e-16
```

So, a = -57.402   Se-a = 3.782

Then, I obtained the b coefficient from the slope for the mediator, top 10, in the model predicting the distal outcome, graduation rate, using both the mediator and exogenous variable, admission rate.

```
lm(formula = Grad.Rate ~ Top10perc + AccRate)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 58.67458    3.68804  15.909   <2e-16 ***
Top10perc    0.45175    0.03456  13.070   <2e-16 ***
AccRate     -7.57898    4.14489  -1.829   0.0679 .

Residual standard error: 14.91 on 774 degrees of freedom
Multiple R-squared:  0.2483,	Adjusted R-squared:  0.2463
F-statistic: 127.8 on 2 and 774 DF,  p-value: < 2.2e-16
```

So, b = 0.45175 Se-b = 0.03456

Thus, the indirect effect estimate, ab, was:

```
ab = -57.402*0.45175 = -25.93135
```

Then, I used Meeker & Escobar's (1994) distribution of products procedure to calculate the 95% coffidence interval estimate of the indirect effect using R's RMediation package and the medic function as follows

```
> library(RMediation)
> #95% confidence interval with Meeker & Escobar's distribution of pro
duct
> medci(mu.x = -57.402, mu.y = 0.45175, se.x = 3.782, se.y =  0.03456,
type = "dop")
$`97.5% CI`
[1] -31.08719 -20.97980
```

(4) Conclusion:

The result led me to reject H0 and infer that the ratio of the applicants accepted and received by a college affects a college's graduation rate through the indirect effect on the ratio of top 10 students enrolled in this college.

[ab =-25.931, distribution of products 95% CI = (-31.087, -20.980), p <0.05]