Friday Quiz 2

School Data

This synthetic dataset simulates data collected from a high school's academic enrichment program, designed to explore how various student habits relate to academic success. In particular, we imagine researchers created three survey-based variables: **x1**, representing weekly hours spent in *peer-led study groups*; **x2**, weekly hours dedicated to *quiet individual study time*; and **x3**, weekly hours spent in *school-sponsored extracurricular clubs*. The outcome variable, **y**, represents the student's final *academic performance index* for the term, scaled from 0 to 100.

The data was generated in such a way that the three predictors are close to statistically independent (i.e., there's no strong correlation between them), allowing for a clean interpretation of each variable's unique contribution to academic performance. For example, a student could study independently without necessarily being involved in peer groups or clubs, making these behaviors distinct. In this synthetic model, we assume that **individual study (x2)** has the strongest positive association with performance, **peer groups (x1)** have a moderate positive effect, and **extracurricular clubs (x3)** show a slight negative association—perhaps reflecting a trade-off in time. The goal of this simulation is to help students and teachers analyze how various independent efforts and activities can affect outcomes when not confounded by overlapping behaviors.

Review of Model Metrics

type	RSS	RSE	R2	Adj_R2	AIC	BIC
Model 1: Performance ~ Study Group	600.84	2.48	0.16	0.15	469.10	476.92
Model 2: Performance ~ Study Group + Self Study	196.11	1.42	0.73	0.72	359.14	369.56
Model 3: Performance ~ Study Group + Self Study + Clubs	106.16	1.05	0.85	0.85	299.76	312.79

Model Comparisons

Model Comparison 1

```
anova(school_model_1, school_model_2)
```

```
Analysis of Variance Table

Model 1: performance ~ study_group
Model 2: performance ~ study_group + self_study
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     98 600.84
2     97 196.11  1    404.72 200.18 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model Comparison 2

```
anova(school_model_2, school_model_3)
```

```
Analysis of Variance Table

Model 1: performance ~ study_group + self_study
Model 2: performance ~ study_group + self_study + clubs
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     97 196.11
2     96 106.16  1    89.955 81.347 1.903e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Review of Selected of Final Model

```
summary(school_model_3)
```

```
Call:
lm(formula = performance ~ study_group + self_study + clubs,
    data = school_df)

Residuals:
     Min       1Q   Median       3Q      Max
-2.50590 -0.65280  0.05288  0.68192  2.52490

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  63.0830     1.0670  59.121  < 2e-16 ***
study_group   0.9453     0.1122   8.424 3.56e-13 ***
self_study    2.0468     0.1095  18.695  < 2e-16 ***
clubs        -1.0532     0.1168  -9.019 1.90e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.052 on 96 degrees of freedom
Multiple R-squared:  0.8519,    Adjusted R-squared:  0.8472
F-statistic:   184 on 3 and 96 DF,  p-value: < 2.2e-16
```