P1. This is an observational study because no human intervention and no control variables exists. Population inferences is not applicable with 21 selected roller coasters (not randomly selected). Causal inferences is not applicable with no random assignment exists.

P2. (a)

(b) The relationship between top speed and height is linear, and strong because there exists an obvious straight line fitting the scattered points, most points falls on the line and few points scatter around. It’s a positive association, with the increase in height, there is an increase in top speed.

(c)

The relationship between top speed and height is likely to be linear, and moderate because points scatter a lot. It’s neither a positive association nor a negative one, with the increase in height, top speed goes up and down (fluctuation exists).

(d) Correlation coefficient top speed vs height: 0.9599

Correlation coefficient top speed vs length: 0.6950

(e) For top speed vs height, 0.9599 represents a significant high linear correlation between the two variables, it is consistent with part (b). For top speed vs length, 0.6950 represent a moderate linear correlation between the two variables, consistent with part (c).

(f) R Square = 0.9214, since the value is pretty close to 1, we claim that the regression model fit the data well.

(g) R Square = 0.4830, since the value is closer to 0 than to 1, we claim that the regression model does not fit the data very well.

P3. (a)



Top Speed ŷ = β0 + β1\*x = 60.8145 + 0.9205 \* Height

(b) Top speed shall have a linear relationship with height. Constant variance, for each value of height, top speed shall has the same standard deviation σ. Top speed values shall be approximately normally distributed and independent.

Significance level: α = 0.05

H­0: β1 ≤ 0; HA: β1 > 0

We apply t-distribution, test statistic t0 = β1 / S.E. = 0.9205/0.0617 = 14.9190

According to t-table, df = 19, 0 < p-value < 0.0005

Reject null hypothesis, there is sufficient evidence that the slope β1 is greater than 0.

(c) Kondaa roller coaster: height = 50m

Predicted ŷ = 60.8145 + 0.9205 \* 50 = 106.8395

Observed y - Predicted ŷ = 113 – 106.8395 = 6.1605

Error = (Observed y - Predicted ŷ) / Observed y = 6.1605 / 113 = 5.4518% < 10%, so good prediction.

(d)

In the residual plot, the residuals are randomly scattered about zero with fairly no systematic increase or decrease in variance and no clear pattern. Not enough evidence that the linear regression assumptions are violated.

P4. (a)



Top Speed ŷ = β0 + β1\*x = 61.0480 + 39.8630 \* Length

(b) Top speed shall have a linear relationship with length. Constant variance, for each value of length, top speed shall has the same standard deviation σ. Top speed values shall be approximately normally distributed and independent.

Significance level: α = 0.05

H­0: β1 ≤ 0; HA: β1 > 0

We apply t-distribution, test statistic t0 = β1 / S.E. = 39.8630/9.4622 = 4.2129

According to t-table, df = 19, 0 < p-value < 0.0005

Reject null hypothesis, there is sufficient evidence that the slope β1 is greater than 0.

(c) Kondaa roller coaster: length = 1.2000m

Predicted ŷ = 61.0480 + 39.8630 \* 1.2000 = 108.8836

Observed y - Predicted ŷ = 113 – 108.8836 = 4.1164

Error = (Observed y - Predicted ŷ) / Observed y = 4.1164 / 113 = 3.6428% < 10%, so good prediction.

(d)

In the residual plot, the residuals are randomly scattered about zero with a likely systematic increase in variance. It is debatable if the constant variance assumption holds.

P5. The percentage of variance (R Square) is higher and closer to zero for the regression of top speed on height than the other one. And based on the assumptions, the Height Residual Plot performs better. Also, the standard error of the regression line parameters is lower for the regression of top speed on height, p-value is lower as well, provides more sufficient evidence that the slope of top speed vs. height is positive.

According to above, the regression model for top speed on height is better. The height variable is more reliable to maximize the top speed due to their significant stronger linearity.