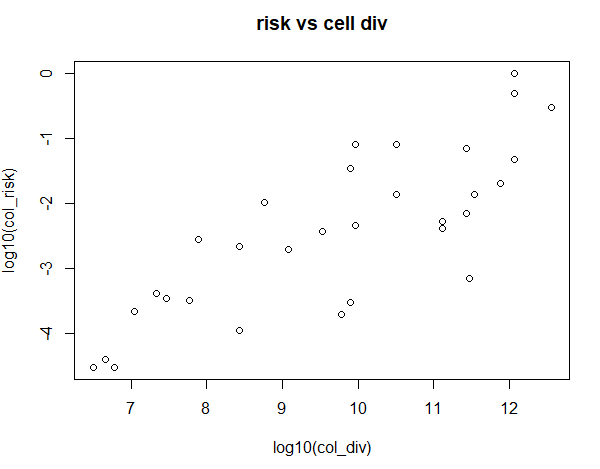
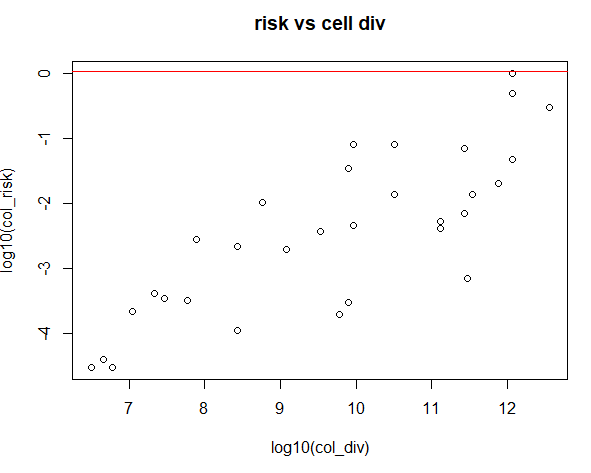
1. **Problem (1A): plot (cancer risk vs cell divisions)**



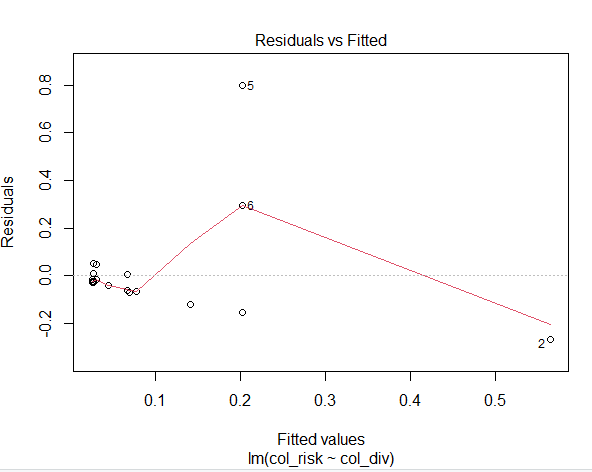
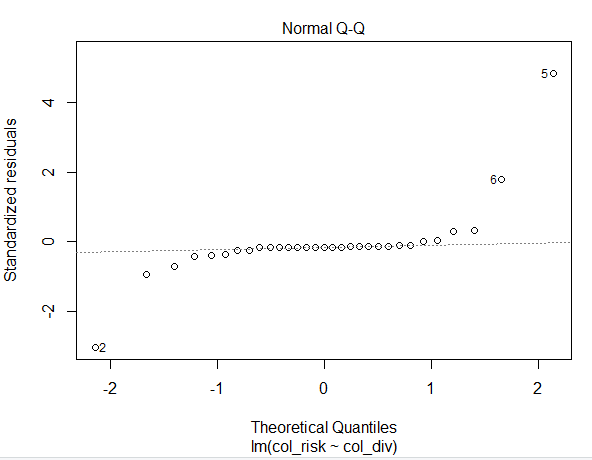
1. **Problem (1B): add regression line**



1. **Problem (1C): p-value, r-squared?**

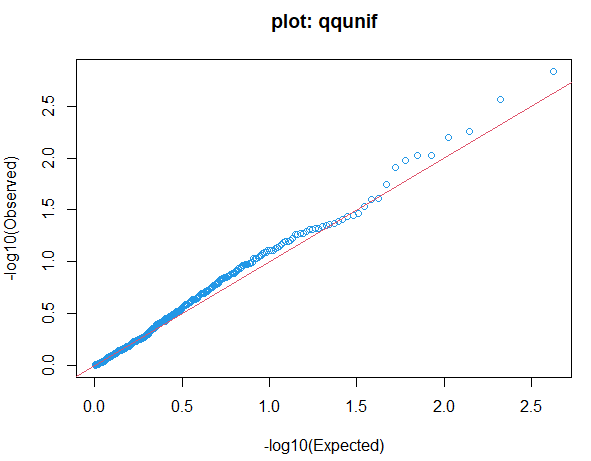
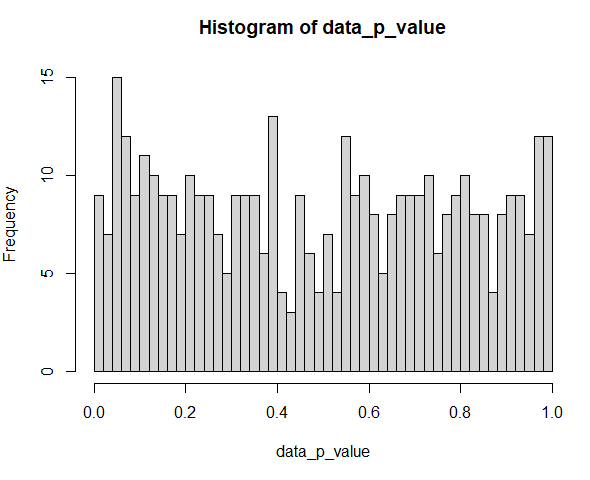
p-value: **0.002027674**, r-squared value: **0.2839264**

1. **Problem (1D): Are the assumptions of constant variance and normal distribution of the residues reasonable for this model? Justify.**

From these two plots, where we can see the **Residuals vs Fitted values** and **Std. Residuals vs Quantiles**, it seems like for most of the samples the residuals are very low indicating that the model almost fits the data based on the underlying assumptions. So, I guess we can say that the assumptions of constant variance and normal distribution of the residues are reasonable for this model (except for very few data points).

1. **Problem (2):**  **Graph out all the p-values. Do they appear uniformly distributed? Does the microbial community appear to be influencing body weight in this cohort?**



I have used 50 breaks to plot the histogram. From the **‘histogram’** and the **‘qqunif’** plots it appeared to me that they are almost uniformly distributed, again these are not perfectly uniform but the overall look seems pretty close.

As a result, I don’t think the microbial community appear to be influencing the body weight much in this cohort, they seem to be almost independent.

**Are any of these associations significant at a 10% false discovery rate?**

For 10% false discovery rate, the unadjusted p-values show 52 significant values out of 420. But when I have used the adjusted ‘BH’ method, none of them were significant (0 out of 420).