PURPOSE The primary goal for this homework assignment is to learn to conduct an indicator linear model analysis. This includes:

- Becoming familiar with the indicator linear model analysis process in statistical terms.
- Learning how to carry out the necessary computations in R.
- Learning how to translate among the “real world” that you are modeling, statistical concepts and terminology, and the R statistical computing environment.

BACKGROUND Use the masses and years of pennies data set (ID #9) from the Data Hoard for this assignment. In 1982, the chemical composition of the penny was changed. In this assignment, you will investigate two related non-statistical questions of interest:

1. Is there a difference in average masses of pennies minted in 1981 or earlier and pennies minted in 1983 or later?

2. What is the difference in average masses of pennies minted in 1981 or earlier and pennies minted in 1983 or later? (A “no” answer to the previous question doesn’t imply an answer for this question because the answer to this question could be 0 grams.)

To translate these into statistical questions, define mass to be the random variable that is the mass of a randomly selected penny, and let era be the 2-level categorical variable defined by:

\[
\text{era} = \begin{cases} 
\text{early} & \text{if the penny was minted in 1981 or earlier} \\
\text{late} & \text{if the penny was minted in 1983 or later.}
\end{cases}
\]

Also, let \( \beta_0 \) denote the coefficient of the constant term in the (unknown) true linear model of class

\[ \text{mass} \sim \text{era} \]

with reference level early for era. Also, let \( \beta_1 \) denote the coefficient of eralate in this true linear model.

With this notation, the non-statistical questions of interest (or really tasks of interest) are:

1. Conduct a significance test of:

\[ H_{null} : \beta_1 = 0 \text{ grams} \quad \text{and} \quad H_{alt} : \beta_1 \neq 0 \text{ grams}, \]

evaluating the result at the usual significance level of 0.05.

2. Compute a 95% confidence interval for \( \beta_1 \).

Before proceeding to the next section, look at the context of the data. Doing so should include reading the description of the data set, taking note of any peculiarities that you find with this, and examining the units for each variable.
**Tasks to accomplish** For each of these tasks, include all relevant R code and output. Be sure to precede each line of R code with an explanation of what it does (or why you are entering it). Also, after each bit of R output, be sure to explain what that R output tells you (or how it helps to accomplish the task at hand).

1. Write the model equation for the true linear model of class mass $\sim$ era with reference level early for era. Also, write the model equation for the fitted model of this class. (These should have $\mu$ and $\beta$ in them. If you can’t type $\hat{\beta}$, feel free to use $\beta^*$ instead, and similarly for $\hat{\mu}$.)

2. (Examine the data.) Examine the sample of mass graphically and comment on anything noteworthy that this reveals (or a lack thereof).

3. (Fit the model.) Fit the model of class mass $\sim$ era. To verify that you have fitted the model, output the fitted model’s coefficients.

4. (Check the SVAs.) Check graphically that the conditional error terms for the levels of era have equal variances.

5. (Check the SVAs.) Check graphically that the conditional error term for each level of era is normally distributed.

6. (Check the SVAs.) Check numerically that the conditional error term for each level of era is normally distributed. Comment on how this fits with what you found graphically.

7. (Draw statistical inferences.) Conduct the required hypothesis test “by hand” in R by computing the test statistic and the requisite area. Then verify your result by reading the result of this hypothesis test from the model summary. Complete the significance test by comparing the $p$-value to the significance level.

8. (Draw statistical inferences.) First compute the required confidence interval “by hand” in R by using the formula for a confidence interval. Then compute the same confidence interval with the `confint()` function.

9. (Interpret your results.) Report what your findings tell you about the statistical and non-statistical questions of interest.