Homework #8

Due: Tuesday, November 23 @ 6pm

Please remember to give R code, as well as answers, for any problems where you used R

Problem 1:

Consider the fungus growth data from HW7:

Laetisaric acid concentration (uG/mL)	0	0	3	3	6	6	10	10	20	20	30	30
Fungus growth (mm)	33.3	31	29.8	27.8	28	29	25.5	23.8	18.3	15.5	11.7	10

a. What is the mean and standard deviation of **both** fungus growth and acid concentration?
b. Using the values from (a) and the correlation coefficient calculated in HW7, calculate the estimates of β_1 and β_0 for the linear model: Growth = β_0 + $\beta_1$concentration. *Show your work*
c. Explain in words what the estimates β_1 and β_0 mean in this dataset.
d. Now, using R's `lm()` function, calculate the estimates of β_1 and β_0 for the linear model: Growth = β_0 + $\beta_1$concentration. How do your answers compare to (b)?
e. Using the equation for your regression line calculated in (b), how much growth (in mm) would you predict the fungus would have when exposed to 15 uG/mL? *Show your work*
f. Now, using R's `predict()` function and the linear model from (d), repeat (e). Is the answer the same?
g. Given the linear model generated for fungus growth and acid concentration, I estimate that at 35 uG/mL of laetisaric acid there will be approximately 6.9 mm of fungal growth. Do you agree with this statement? Why or why not?

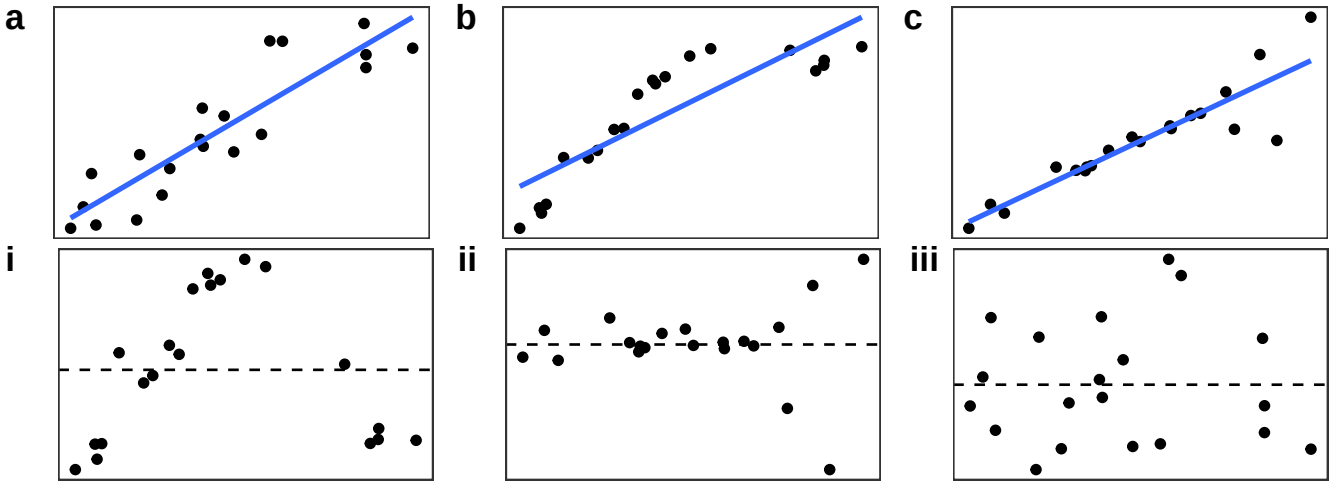
Problem 2:

Let's continue with the fungus data from **Problem 1**

a. Calculate the `SS(resid)` or residual sum of squares for the linear model you derived in **Problem 1**. There are many ways to do this, but the simplest is probably to use `deviance(model)` where `model` is the linear model output from `lm()`.
b. Describe in words what the `SS(resid)` is measuring. (*Note: "residual sum of squares" is not an appropriate answer*)
c. Calculate the coefficient of determination, r^2. Explain in words what this value means in this context.
d. Consider the null hypothesis that laetisaric acid has no effect on growth of the fungus. Assuming that the linear model is applicable, state in symbols the null hypothesis about the true regression line and an alternative hypothesis that laetisaric acid inhibits growth of the fungus.
e. How many degrees of freedom are there for the test in (d)?
f. Calculate the test statistic for the test in (d) using `summary(model)` where `model` is the linear model output from `lm()`. **Be sure to write out the test statistic, just showing the model output is not good enough**
g. Using the `*t()` functions or the output from (f), state your conclusion regarding the null hypothesis in this context (*Note: remember to consider your alternative hypothesis). Be sure to provide the p-value, alpha, whether you reject or not, and what your final conclusion is.

Problem 3:

a. Describe the conditions/assumptions for using a linear model
b. The following three residual plots (i), (ii), and (iii) were generated after fitting regression lines to the following three scatterplots (a), (b), and (c). Which residual plots goes with which scatterplot? **How do you know?**



c. Plot the residuals from the fungus data in **Problem 1**. Which condition/assumption are we checking with a residual plot? Is it met for the fungus data?

Problem 4:

The following dataset contains information about newborn babies and their parents. There is a lot of information here, but suppose we are interested in determining if the number of weeks at which a baby is born (i.e. `Gestation` (weeks)) can predict baby weight (i.e. `Birthweight` (kg)).

```
url <- "https://raw.githubusercontent.com/katiesevans/IGP_biostatistics/main/data/birthweight.csv"
babies <- read.csv(url)
```

a. Read in the data with the code above and plot gestation versus birth weight. Describe what you see. Does there appear to be a linear relationship between the variables? What is the correlation coefficient?
b. Assuming all assumptions are met, use R to generate a linear model of birth weight given gestational age. Add the regression line to your plot.
c. Write a few sentences detailing your results **as you would in a manuscript**. Make sure to include:

- What the question was
- What test was performed
- The equation of the linear model
- The variance explained
- Written explanation of the regression coefficient
- Confidence interval for the regression coefficient
- p-value of the test
- final conclusion (including whether or not you can claim causality)