Module 3E Questions:

1. From the 12-mouse dataset: **Test the following hypothesis**: Are mice with greater weight gain more likely to have higher glucose? Would you expect the correlation to be positive, negative, or near zero?

2. Scientists find that **activity level** is correlated with **PPARG expression** across a group of 20 mice. They now ask whether **PPARG expression can *predict* activity level** in new mice.
   - Explain conceptually why **correlation** alone cannot answer this prediction question, and why **regression** is the appropriate tool when the goal is *prediction*.
   - In your explanation, describe how the two methods differ in terms of:

     - what they measure,
     - whether they assign predictor/outcome roles, and
     - what kinds of questions they are designed to answer.

3. Researchers want to explore the relationship between gene length (in kilobases) and gene expression (in TPM) in a sample of 15 genes from a particular mouse strain. *Is there a significant relationship between gene length and gene expression?*

Characteristic	Mean	Standard Deviation
Gene Length	8.2 kb	1.5kb
Gene Expression	120 TPM	15 TPM

   The correlation coefficient is r=0.65. Interpret this value and test the significance of the correlation. You do not need to provide a confidence interval!