**INL2: Data Analysis**

**Scenario 1**

**Answer the following questions**

**1. What are the objects, subjects, treatment and factors in this experiment?**   
In this experiment, the objects are the programming tasks Prog-1 and Prog-2, which the participants are required to implement. The subjects are the 10 programmers, each of whom completes both tasks. The treatment is the use of one of the two IDEs (IDE-A or IDE-B) during development. The factor is the type of IDE, which is the variable we control to assess its impact on efficiency.

**2. How would you describe this in terms of a standard design type e.g. one factor, two treatments?**   
The experiment represents a design with one factor (IDE type) and two treatment levels (IDE-A and IDE-B). It is a balanced crossover design, where each participant completes both tasks but in a different order, allowing us to control for individual differences and order effects.

**3. In Table 2, why are five programmers using IDE-A to develop program Prog-1 first and then developing Prog-2 using IDE-B, while the remaining five programmers will use IDE-A to develop program Prog-2 first and then use IDE-B to program Prog-1? What is the benefit of this design?**   
This approach helps eliminate systematic errors related to task order. If all participants started with the same IDE, the results could be skewed due to learning effects or fatigue. The crossover distribution ensures balance and makes the comparison more reliable.

**4. Why is it important to assume that Prog-1 and Prog-2 are different, but still equally complex/hard to develop? What are the implications on the design and conclusions of the study if this is not true?**   
If one task is objectively easier or harder, it may affect completion time regardless of the IDE used. In that case, time differences would be due to the task itself, not the IDE, leading to unreliable conclusions. Assuming equal complexity is essential to maintain internal validity of the study.

**Answer the following questions related to analysis**

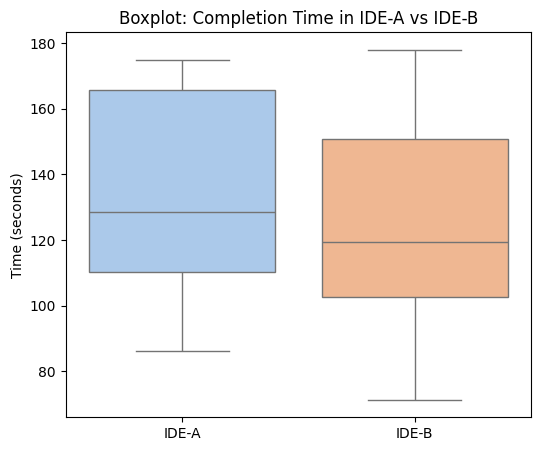
**1. State the null and alternative hypothesis for this investigation.**

(H₀): A =B. There is no significant difference in task execution time between IDE-A and IDE-B.

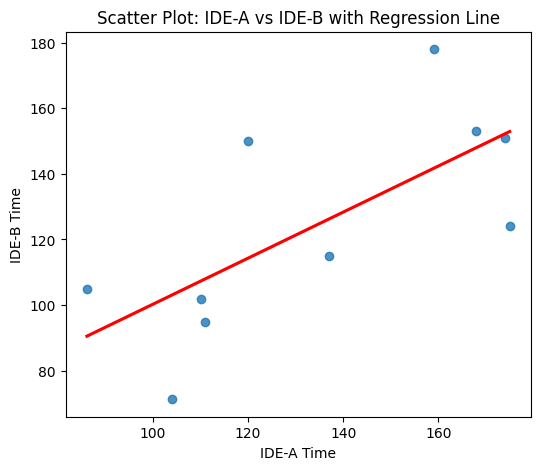
(H₁): A B There is a significant difference in task completion time between IDE-A and IDE-B.

**2. Use descriptive statistics and visualize the data. Which visualization tool helped you develop some insights into the data? What were the insights e.g. interesting patterns or trends in the data, a clear difference in efficiency between two IDEs, outliers. How will you group the data for these visualizations?**   
For the analysis, I grouped the data based on task completion time depending on which IDE was used by each programmer. The average time for Prog-1 was 131.10 minutes, for Prog-2 — 127.73 minutes. The average difference between IDE-A and IDE-B was 9.99 minutes, with a standard deviation of 25.50 minutes.  
For visualization, I used the following plots:

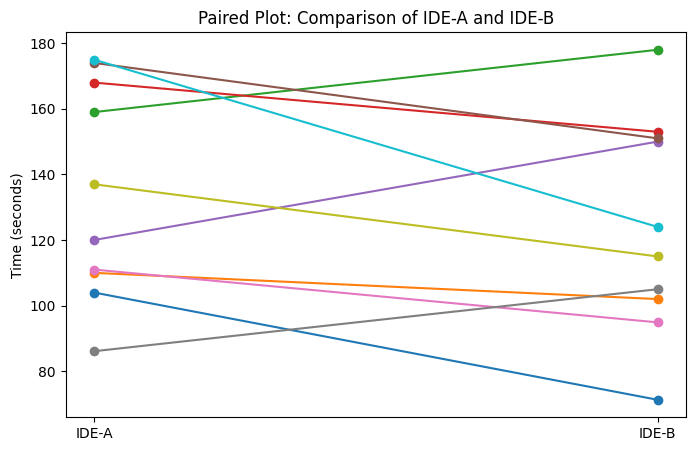
**Boxplot** - showed that the medians of IDE-A and IDE-B are close, but IDE-B has slightly more spread.

**

**Scatter plot with regression line** - revealed a weak correlation between IDE-A and IDE-B times.

**

**Paired plot** - clearly illustrated individual differences between IDE-A and IDE-B.

**

**3. Choose and justify your choice of a parametric/non-parametric method for analyzing the given data. Document the steps you undertook and the results.**   
I performed normality tests on the distribution of time differences between IDE-A and IDE-B.

* Shapiro-Wilk Test: W=0.941, p=0.569
* Kolmogorov-Smirnov Test: D=0.178, p=0.857

Both tests showed that the data does not deviate from normal distribution, so I chose a parametric method - the paired t-test, which is suitable for dependent samples and normally distributed data.

**4. Run the statistical method and report if you can reject the null hypothesis. Please interpret your results, what does this imply for the objective of the study?**   
Results of the paired t-test: t = 1.239, p = 0.247  
Since p > 0.05, we do not reject the null hypothesis. This means that no statistically significant difference was found between IDE-A and IDE-B. Therefore, we cannot conclude that one IDE provides better efficiency in terms of task completion time.

**5. Based on the results would you be confident to recommend an IDE either IDE-A or IDE-B for use in your company. Why or why not?**   
Based on the results, I would not confidently recommend either IDE. Although IDE-B showed slightly lower average completion time, the difference is not statistically significant. Moreover, the spread of values and individual differences among programmers suggest that IDE efficiency may depend on personal preferences and experience. A more extensive study with more participants and tasks would be needed for a well-founded recommendation.