Project 1 for Project Check-ins #2

Group Information

- Group Number: 15
- Group Members: Siwoong Yoon, Minsun Jee, Seungwoo Kim, Jeonghyun Kim, Yewon Jeung
- Date of Discussion: 2025. 10. 07

Part A

1. Regression Model Specification

Model: [Purchase amount(Y)] = β_0 + β_1[Age(X)]

Expected Interpretation:

- **β_0:** The expected baseline purchase amount when age is zero. (Since this value is outside the data's age range, it should be interpreted as a mathematical intercept.)
- **β_1:** The change in average purchase amount per one-year increase in age.

Business Translation:

This model informs us how much the custom's average purchase amount changes for every one year increase in age. Through this analysis, we can determine if there is a statistically significant linear relationship between age and purchase amount, and how well age alone can explain the variability in purchase amount.

2. Comparison Table: Testing vs. Regression

Analysis Component	W3 Hypothesis Test	W4 Regression
Research Question	Is there a statistically significant difference in online shopping behavior(purchase amount) between middle-aged and young people?	How much does age affect the purchase amount?
Null hypothesis	Middle-aged people use online shopping less than young people.	β_1=0. No linear effect of age on purchase amount.
Alternative	Middle-aged people use online shopping the same or more than younger generations.	$\beta_1 \neq 0$. Age has a linear effect on purchase amount.
Decision Metric	p-value(<0.05)	$R^2$and slope magnitude
Expected Finding	difference likely small/unclear	Slightly negative β_1, $R^2 \approx 0.3$ (Radio-like pattern)
Limitations	Hard to generalize;missing income, race, etc.	Same with W3 limitations.

3. R^2 Interpretation Planning

We expect the R^2 is in the 0.10–0.30 range, and it will show that the result has a weak but possibly meaningful relationship. Age is likely related to online shopping behavior, but it's not the only factor that influences spending. Other variables such as income or digital literacy will probably explain more of the variations.

4. Decision Framework

If p-value < 0.05 AND R^2 > 0.30:

→ There is strong evidence that age and purchase amount are significantly related, and the model has moderate predictive power.
→ This means age plays an important role in explaining online shopping behavior, even though other factors may also contribute.

If p-value < 0.05 BUT R^2 < 0.30:

→ The relationship between age and purchase amount is statistically significant but weak in prediction.
→ This suggests that age has an effect, but it explains only a small portion of total variation — similar to the Newspaper case in Lab 1.
→ We can interpret that while age matters, most differences in shopping behavior come from other variables such as income or gender.

If p-value > 0.05:

→ There is no statistically significant linear relationship between age and purchase amount.
→ In this case, we may need to revisit our Week 3 hypothesis testing approach or include other variables that better explain shopping behavior.

Part B

1. Contribution Tracking

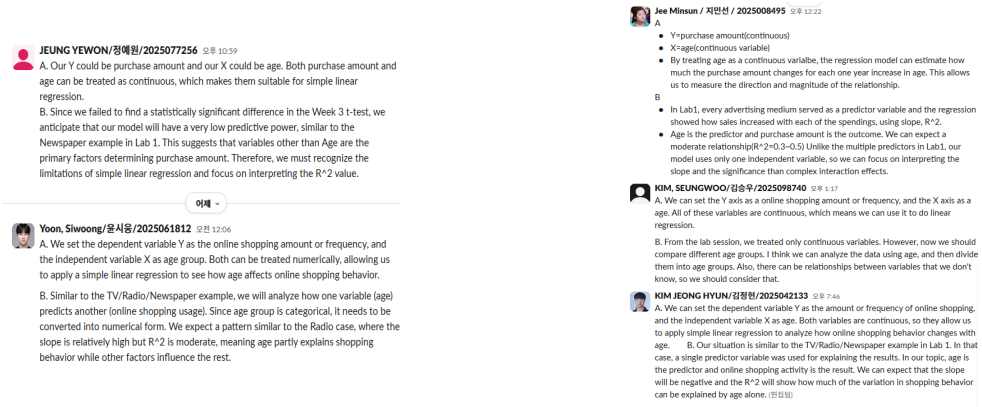

Fig 1. Screenshot of Part of Our Group's Discussion on Slack

Name	Role	Specific section you contributed to
Minsun Jee	Idea giver & Critical thinker	Defining regression variables. Explaining the relationship between age and purchase amount.
Jeonghyun Kim	writer & planner	I wrote about the relationship between R^2 and variables such as age and expected degree.
Yewon Jeung	brainstormer & writer	I wrote the section about regression model specification, by checking other students' opinions.
Seungwoo Kim	Visualization & Writer	I conducted linear regression using our data(this is on our group slack channel) and thought about p-value's meaning.
Siwoong Yoon	Critical Thinker	I clearly determined the final variables to be used in our topic, predicted the expected R^2 value, and suggested the multiple regression model.

2. Evolution Reflection

Our research question has evolved. In the first week, we just focused on a single variable. Our original hypothesis is 'Middle aged people use online shopping less than young generations.'. We continued with this question during the project.

In week 4, however, we faced challenges that we can't find significant relationships in our hypothesis using simple linear regression. Simple linear regression can't derive a significant relationship between purchase amount and age. To deal with this situation, we suggested considering various variables together with multiple regression models. For example, we can consider that older people with higher income and stronger digital skills might shop online as frequently as younger users.