

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.	$\beta_1=0$. No linear effect of age on purchase amount.
Alternative	Middle-aged people use online shopping the same or more than younger generations.	$\beta_1\neq0$. 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 , $R^2\approx0.3$ (Radio-like pattern)
Limitations	Hard to generalize;missing income, race, etc.	Same with W3 limitations.

3. R² Interpretation Planning

We expect the R² 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 R² > 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 R² < 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

JEUNG YEWON/정예원/2025077256 오후 10:59

A. Our Y could be purchase amount and our X could be age. Both purchase amount and age can be treated as continuous, which makes them suitable for simple linear regression.

B. Since we failed to find a statistically significant difference in the Week 3 t-test, we anticipate that our model will have a very low predictive power, similar to the Newspaper example in Lab 1. This suggests that variables other than Age are the primary factors determining purchase amount. Therefore, we must recognize the limitations of simple linear regression and focus on interpreting the R^2 value.

이제 ~

Yoon, Siwoong/윤서웅/2025041812 오후 12:04

A. We set the dependent variable Y as the online shopping amount or frequency, and the independent variable X as age group. Both can be treated numerically, allowing us to apply a simple linear regression to see how age affects online shopping behavior.

B. Similar to the TV/Radio/Newspaper example, we will analyze how one variable (age) predicts another (online shopping usage). Since age group is categorical, it needs to be converted into numerical form. We expect a pattern similar to the Radio case, where the slope is relatively high but R^2 is moderate, meaning age partly explains shopping behavior while other factors influence the rest.

Jee Minsun / 지민선 / 2025008495 오후 12:22

A.

- Y=purchase amount(continuous)
- X=age(continuous variable)
- By treating age as a continuous variable, the regression model can estimate how much the purchase amount changes for each one year increase in age. This allows us to measure the direction and magnitude of the relationship.

B.

- In Lab1, every advertising medium served as a predictor variable and the regression showed how sales increased with each of the spendings, using slope, R^2 .
- Age is the predictor and purchase amount is the outcome. We can expect a moderate relationship($R^2=0.3-0.5$). Unlike the multiple predictors in Lab1, our model uses only one independent variable, so we can focus on interpreting the slope and the significance than complex interaction effects.

KIM, SEUNGWOO/김승우/2025098740 오후 1:17

A. We can set the Y axis as an online shopping amount or frequency, and the X axis as age. All of these variables are continuous, which means we can use it to do linear regression.

B. From the lab session, we treated only continuous variables. However, now we should compare different age groups. I think we can analyze the data using age, and then divide them into age groups. Also, there can be relationships between variables that we don't know, so we should consider that.

KIM JEONG HYUN/김정현/2025042133 오후 7:40

A. We can set the dependent variable Y as the amount or frequency of online shopping, and the independent variable X as age. Both variables are continuous, so they allow us to apply simple linear regression to analyze how online shopping behavior changes with age.

B. Our situation is similar to the TV/Radio/Newspaper example in Lab 1. In that case, a single predictor variable was used for explaining the results. In our topic, age is the predictor and online shopping activity is the result. We can expect that the slope will be negative and the R^2 will show how much of the variation in shopping behavior can be explained by age alone. (전진진)

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.