ADS4293 LAB 5

Chapter 8: More Topics in Regression Analysis

TA: Malak Al-Hanafi

October 9th, 2025

LAB SETUP

- Excel file ch8 questions
- We’ll go over one question at a time
- Submit your **Excel file** on *Brightspace (Owl)* at the end of the lab
- Today’s lab has 4 questions

QUESTION 1 - 8.12

12. FILE ***Pick_Errors.*** The distribution center for an online retailer has been experiencing quite a few "pick errors" (i.e., retrieving the wrong item). Although the warehouse manager thinks most errors are due to inexperienced workers, she believes that a training program also may help to reduce them. Before sending all employees to training, she examines data from a pilot study of 30 employees. The accompanying data file includes information on the employee's annual pick errors (Errors), experience (Exper in years), and whether or not the employee attended training (Train equals 1 if the employee attended training, 0 otherwise).

    a. Estimate two models:

    $\text{Errors} = \beta_0 + \beta_1\text{Exper} + \beta_2\text{Train} + \varepsilon$, and

    $\text{Errors} = \beta_0 + \beta_1\text{Exper} + \beta_2\text{Train} + \beta_3\,(\text{Exper} \times \text{Train}) + \varepsilon$.

    b. Which model provides a better fit in terms of adjusted R^2 and the significance of the predictor variables at the 10% level?

    c. Use the chosen model to predict the number of pick errors for an employee with 10 years of experience who attended the training program, and for an employee with 20 years of experience who did not attend the training program.

    d. Give a practical interpretation for the positive interaction coefficient.

Q 1 A

a. Estimate two models:

$$\text{Errors} = \beta_0 + \beta_1 \text{Exper} + \beta_2 \text{Train} + \varepsilon, \text{ and}$$

$$\text{Errors} = \beta_0 + \beta_1 \text{Exper} + \beta_2 \text{Train} + \beta_3 (\text{Exper} \times \text{Train}) + \varepsilon.$$

- Create interaction column for Exper×Train
- Data → Data Analysis → Regression
- Y Range: Errors
- X Range: Exper, Train, Exper×Train

- For the second model, the **interaction variable** Exper×Train allows the slope of Experience to change for *trained vs. untrained* workers

QUESTION 1B

b. Which model provides a better fit in terms of adjusted R^2 and the significance of the predictor variables at the 10% level?

- If adjusted R^2 is greater, it explains more variation → better fit

- For Model 1: check if Exper and Train are significant
- For Model 2: check if Exper, Train, and Exper×Train are significant
  - If $p < 0.10$, then variable is significant

QUESTION 1C

c. Use the chosen model to predict the number of pick errors for an employee with 10 years of experience who attended the training program, and for an employee with 20 years of experience who did not attend the training program.

QUESTION 1D

d. Give a practical interpretation for the positive interaction coefficient.

- Look at your results from c and think about who benefits more from the **training**?

QUESTION 2- 8.19

19. FILE ***Rental.*** Jonah has worked as an agent for rental properties in Cincinnati, Ohio, for almost 10 years. He understands that the premium for additional bedrooms is more in larger as compared to smaller homes. The accompanying file includes the monthly data on rent (Rent, in $) along with square footage (Sqft), number of bedrooms (Bed), and number of bathrooms (Bath).

    a. Estimate a regression model for monthly rent on the basis of Bed, Bath, Sqft, and the interaction between Bed and Sqft. Interpret the estimated coefficient of the interaction variable.

    b. Predict the monthly rent of homes with a square footage of 1,600, 2 baths, and the number of bedrooms equal to 2, 3, and 4. Compute the incremental predicted rent as the number of bedrooms increases from 2 to 3 and 3 to 4.

QUESTION 2A

a. Estimate a regression model for monthly rent on the basis of Bed, Bath, Sqft, and the interaction between Bed and Sqft. Interpret the estimated coefficient of the interaction variable.

- Create interaction column for Sqft×Bed
- Data → Data Analysis → Regression
- Y Range: Rent
- X Range: Sqft, Bed, Bath, Bed×Sqft
- Manually write the equation based on the coefficient values you get
- Rent = ...

Interpretation:
Is each variable significant or not? (look at p<0.05->Significant → interaction improves fit)

QUESTION 2B

b. Predict the monthly rent of homes with a square footage of 1,600, 2 baths, and the number of bedrooms equal to 2, 3, and 4. Compute the incremental predicted rent as the number of bedrooms increases from 2 to 3 and 3 to 4.

QUESTION 3- 8.36

36. FILE ***Happiness.*** Researchers often rely on surveys to capture a subjective assessment of well-being. One study finds that holding everything else constant, people seem to be least happy when they are in their 40s (*Psychology Today,* April 27, 2018). Another study suggests that money does buy happiness, but its effect diminishes as incomes rise above $75,000 a year (*Money Magazine,* February 14, 2018). The accompanying file includes survey data of 100 working adults' self-assessed happiness on a scale of 0 to 100, along with their age and annual income.

a. Estimate and interpret a regression model for Happiness based on Age, Age^2, and ln(Income).

b. Predict happiness with Income equal to $80,000 and Age equal to 30, 45, and 60 years.

QUESTION 3A

a. Estimate and interpret a regression model for Happiness based on Age, Age^2, and ln(Income).

- Add 2 new columns (Age squared and ln(income))

- Data → Data Analysis → Regression
- Y Range: Happiness
- X Range: select Age, Age^2, ln(Income)

Interpret:

- write the equation with coefficients
- are the coefficients significant at the 5%? <0.05

QUESTION 3B

b. Predict happiness with Income equal to $80,000 and Age equal to 30, 45, and 60 years.

QUESTION 4 - 8.37

37. FILE ***Production_Function.*** Economists often examine the relationship between the inputs of a production function and the resulting output. A common way of modeling this relationship is referred to as the Cobb-Douglas production function. This function can be expressed as $\ln(Q) = \beta_0 + \beta_1 \ln(L) + \beta_2 \ln(K) + \varepsilon$, where Q stands for output, L for labor, and K for capital. The accompanying file contains relevant data.

    a. Estimate $\ln(Q) = \beta_0 + \beta_1 \ln(L) + \beta_2 \ln(K) + \varepsilon$. What is the predicted change in output if labor increases by 1%, holding capital constant?

    b. Holding capital constant, can we conclude at the 5% level that a 1% increase in labor will increase the output by more than 0.5%?

QUESTION 4A

a. Estimate $\ln(Q) = \beta_0 + \beta_1 \ln(L) + \beta_2 \ln(K) + \varepsilon$. What is the predicted change in output if labor increases by 1%, holding capital constant?

- In a **log–log model**, we treat changes as **percentages** because taking logs converts differences into proportional (percent) differences.

- Add 3 new columns lnQ, lnL, lnK
- Data → Data Analysis → Regression
- Y Range: lnQ
- X Range: lnL, lnK

- Each coefficient *(lnL and lnK)* tells you how Q changes (%) for a 1% change in that input, keeping the other constant
- Predicted % change in Q = coefficient × % change in input
  - Basically predicted change will just be the coefficient of ln(L) for a 1%

QUESTION 4B

b. Holding capital constant, can we conclude at the 5% level that a 1% increase in labor will increase the output by more than 0.5%?

- Hypotheses
  - H0: $\beta(L) \leq 0.5$ **Labor elasticity is 0.5 or smaller**
  - HA: $\beta(L) > 0.5$ **Labor elasticity is greater than 0.5**

- We'll use a t-test to decide

$$T_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{se(\hat{\beta}_1)}$$

- $\beta 1$ = our sample slope ln(l)
- $\beta 1,0$ = the hypothesized value (0.5)
  - se($\beta 1$) = the standard error of the slope $\beta 1$

- **One tailed test** (you use a one-tailed test when the alternative uses *> or <*):
  - Find p-value using **=1-T.DIST(x, df, TRUE)**, where x is the t-statistic, and **df** is degrees of freedom [**n-k-1**, n is sample size, and k is number of estimated parameters (48-2-1=**45**)], TRUE bc it gives the cumulative probability
    - IF $p < 0.05$ → Reject H_0✅
    - ELSE → Fail to reject H_0

Thank you!