ADS4293 LAB4

Chapter 7: Regression Analysis

TA: Malak Al-Hanafi

October 2nd, 2025

LAB SETUP

- Excel file ch7 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

38. FILE ***Yields.*** While the Federal Reserve controls short-term interest rates, long-term interest rates essentially depend on supply/demand dynamics, as well as longer-term interest rate expectations. In order to examine the relationship between short-term and long-term interest rates, monthly data were collected on the three-month Treasury yield (Three_month, in %) and the 10-year Treasury yield (Ten_year, in %). The accompanying data file shows the results.

    a. Estimate and interpret a sample regression equation using the 10-year yield as the response variable and the three-month yield as the predictor variable.

    b. Interpret the coefficient of determination.

    c. At the 5% significance level, is the three-month yield significant in explaining the 10-year yield?

    d. Many wonder whether a change in the three-month yield implies the same change in the 10-year yield. At the 5% significance level, is this belief supported by the data?

Q 1 A

a. Estimate and interpret a sample regression equation using the 10-year yield as the response variable and the three-month yield as the predictor variable.

We want to see if short-term (3-month) rates explain long-term (10-year) rates
Use simple linear regression:

$$\textit{Ten_year} = b0 + b1 * \textit{Three_month} + \varepsilon$$

1. Data → Data Analysis → Regression
2. Y Range: Ten_year column
3. X Range: Three_month column
4. Include headers and Check Labels

Interpreting the results:

- Look at Coefficients table:
  - Intercept → b0
  - Three_month (slope) → b1

Q 1 B

b. Interpret the coefficient of determination.

R^2 → tells us how much variation in Y is explained by X

- A higher R^2 means the model explains more of the variation in the dependent variable (better fit).
- A lower R^2 means the model explains less of the variation (weaker fit).

Q 1 C

c. At the 5% significance level, is the three-month yield significant in explaining the 10-year yield?

- The null hypothesis (H_0) says: “there’s no relationship, the slope β1 is zero.”
- The alternative (H_A) says: “there is a relationship, the slope β1 isn’t zero.”

Look at the **p-value** for the slope (Three_month)

- If P < 0.05 → slope is statistically significant (reject the null hypothesis H0:β1=0).
- If P ≥ 0.05 → not significant at 5%.

Q 1 D

d. Many wonder whether a change in the three-month yield implies the same change in the 10-year yield. At the 5% significance level, is this belief supported by the data?

- Null hypothesis (H_0): β1=1 (slope equals 1 → one-for-one change)
- Alternative hypothesis (H_A): β1≠1 (slope not equal to 1)

1. We know the estimated slope from the regression is **0.6207**, which is less than 1
2. Now, we calculate how far away 0.6207 is from 1, measured in "standard errors"

We'll use a **t-test** to decide:

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

- β1 = our sample slope
- β1,0 = the hypothesized slope under the null (what we're testing against)
    - 0 for "no effect," or 1 for "one-for-one effect"
- se(β1) = the *standard error* of the slope β1 (how much uncertainty there is in that estimate

3. Find p-value using **=T.DIST.2T(x, df)**, where x is the t-statistic, and **df** is degrees of freedom [**n-k**, n is sample size, and k is number of estimated parameters (intercept β0 + slope β1 = 2)]

- If the p-value > 0.05, then we **fail to reject** H_0 → slope = 1 (one-for-one change could be true)
- If the p-value < 0.05, then we **reject** H_0 → slope ≠ 1 (one-for-one change is not true)

QUESTION 2

41. **FILE** ***Ice_Cream.*** A manager at an ice cream store is trying to determine how many customers to expect on any given day. Overall business has been relatively steady over the past several years, but the customer count seems to have ups and downs. The accompanying data file shows the number of customers, the high temperature (Temperature in degrees Fahrenheit), and whether the day fell on a weekend (Weekend equals 1 if weekend, 0 otherwise) over the past 30 days.

    a. Estimate Customers = $\beta_0 + \beta_1$Temperature + β_2Weekend + ε.

    b. How many customers should the manager expect on a Sunday with a forecasted high temperature of 80°?

    c. Interpret the estimated coefficient for Weekend. Is it significant at the 5% level? How might this affect the store's staffing needs?

QUESTION 2A

a. Estimate Customers = $\beta_0 + \beta_1$Temperature + β_2Weekend + ε.

- Data → Data Analysis → Regression
- Y Range: Customers column
- X Range: both Temperature and Weekend columns
- Include headers and Check Labels

QUESTION 2B

b. How many customers should the manager expect on a Sunday with a forecasted high temperature of 80°?

- use the forumla

$$\text{Customers} = \beta_0 + \beta_1 \text{Temperature} + \beta_2 \text{Weekend}$$

QUESTION 2B

c. Interpret the estimated coefficient for Weekend. Is it significant at the 5% level? How might this affect the store's staffing needs?

In the coefficients table row Weekend:

- Coefficient ≈ ??
- P-value ≈ 3.8E−13 ≪ 0.05 (approx. 0)
  - If p-value < 0.05 → predictor is significant
  - If p-value > 0.05 → predictor is not significant

- On weekends, the store gets **???** more customers than on weekdays at the same temperature.
- Is it statistically significant at 5% ?
- Staffing: plan more or less staff on weekends?

QUESTION 3

46. FILE ***Retail.*** A government researcher is analyzing the relationship between retail sales (in $ millions) and the gross national product (GNP in $ billions). He also wonders whether there are significant differences in retail sales related to the quarters of the year. The accompanying data file shows information on 10 years of quarterly data.

a. Estimate $y = \beta_0 + \beta_1 x + \beta_2 d_1 + \beta_3 d_2 + \beta_4 d_3 + \varepsilon$, where y and x represent retail sales and GNP, respectively; d_1 equals 1 if quarter 1, 0 otherwise; d_2 equals 1 if quarter 2, 0 otherwise; and d_3 equals 1 if quarter 3, 0 otherwise. Note that the reference category is quarter 4.

▸ **Answer**

b. Predict retail sales in quarters 2 and 4 if GNP equals $20,000 billion.

▸ **Answer**

c. Which of the quarterly sales are significantly different from those of the 4th quarter at the 5% level?

▸ **Answer**

d. Reformulate the model to determine, at the 5% significance level, if sales differ between quarter 2 and quarter 3. Your model must account for all quarters.

QUESTION 3A

a. Estimate $y = \beta_0 + \beta_1 x + \beta_2 d_1 + \beta_3 d_2 + \beta_4 d_3 + \varepsilon$, where y and x represent retail sales and GNP, respectively; d_1 equals 1 if quarter 1, 0 otherwise; d_2 equals 1 if quarter 2, 0 otherwise; and d_3 equals 1 if quarter 3, 0 otherwise. Note that the reference category is quarter 4.

Make 3 coloumns for dummy variables d1,d2,d3 and use **if** statements for each

- d1 = 1 if Quarter=1 else 0
- d2 = 1 if Quarter=2 else 0
- d3 = 1 if Quarter=3 else 0
- Q4 is reference

1. Data → Data Analysis → Regression
2. Input Y Range: Sales
3. Input X Range: GNP, d1, d2, d3
4. Check Labels → choose an Output Range → OK

QUESTION 3B

b. Predict retail sales in quarters 2 and 4 if GNP equals $20,000 billion.

- Use equation $y = \beta_0 + \beta_1 x + \beta_2 d_1 + \beta_3 d_2 + \beta_4 d_3 + \varepsilon$
- for Q2, set $d2=1$
- For Q4 (reference): all dummies 0

QUESTION 3C

c. Which of the quarterly sales are significantly different from those of the 4th quarter at the 5% level?

- In the Coefficients table, check p-values for d1, d2, d3.
- $p < 0.05$ → that quarter's mean differs from Q4 (at the same GNP)
- $p \geq 0.05$ → no evidence of a difference vs Q4.

QUESTION 3D

d. Reformulate the model to determine, at the 5% significance level, if sales differ between quarter 2 and quarter 3. Your model must account for all quarters.

- To test Q2 vs Q3, you only need to make one of them the reference.
- Whether you choose Q3 as the reference (compare Q2 against Q3), or Q2 as the reference (compare Q3 against Q2), the conclusion is identical.
  - d1 = 1 if Q1 else 0
  - d3 = 1 if Q3 else 0
  - d4 = 1 if Q4 else 0
  - **Q2 is the reference**
- Run the same regression with GNP, q1, q3, q4
- If p-value < 0.05, then Q2 ≠ Q4.
- If p-value ≥ 0.05, then Q3 ≈ Q4 (no significant difference).

Thank you!