ADS4293 LAB3

Chapter 4 - Data Visualization, Chapter 7: Correlation & Simple Regression

TA: Malak Al-Hanafi

25 September, 2025

LAB SETUP

- Excel file for ch4 & 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 2 questions

QUESTION 4

35. FILE ***Healthy_Living.*** Healthy living has always been an important goal for any society. Most would agree that a diet that is rich in fruits and vegetables (FV) and regular exercise have a positive effect on health, while smoking has a negative effect on health. The accompanying data file shows the percentage of these variables observed in various states in the United States.

    a. Construct a scatterplot of Health against Exercise. Describe the relationship.

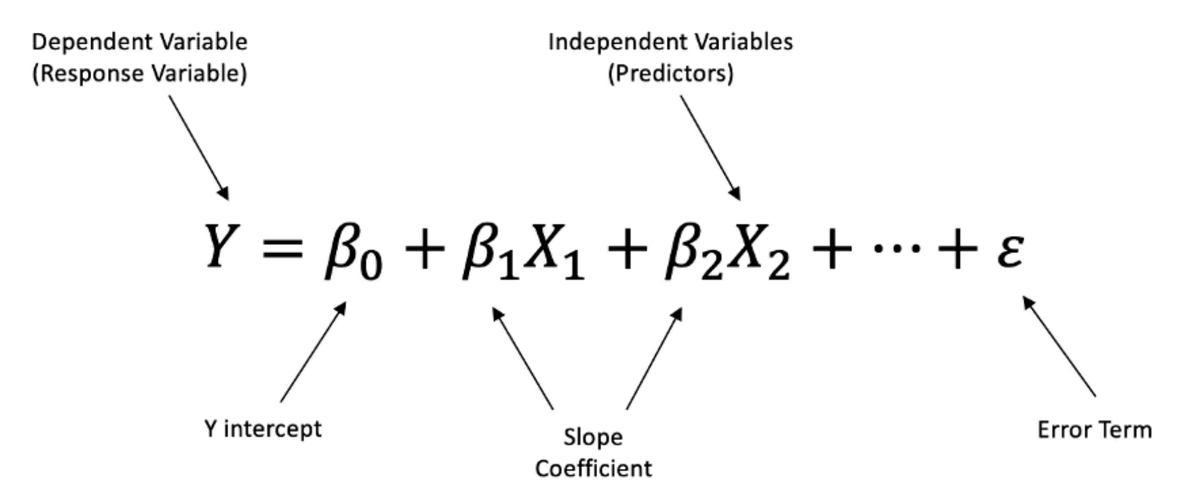
    b. Construct a scatterplot of Health against Smoking. Describe the relationship.

INTRODUCTION TO REGRESSION

What is regression?

A statistical method that helps us predict one variable (the outcome)
Based on the values of other variables (the predictors)

Think of it like a line (or plane) of “best fit” through the data



Examples

- Predicting **profit** from advertising *investment* and *demand*
- Predicting **health** score from *exercise* and *smoking*

CH.7 - QUESTION 1

17. FILE ***Quarterbacks.*** American football is the highest-paying sport on a per-game basis. Given that the quarterback is considered the most important player on the team, he is typically well compensated. A sports statistician examines the factors that influence a quarterback's salary. He believes that a quarterback's pass completion rate is the most important variable affecting salary. He also wonders how total touchdowns scored and a quarterback's age might impact salary. The accompanying data file shows information that he has collected on salary (Salary in $ millions), pass completion rate (PC in %), total touchdowns scored (TD), and age for 32 quarterbacks during a recent season.

a. Estimate: Salary = $\beta_0 + \beta_1 PC + \beta_2 TD + \beta_3 Age + \varepsilon$.

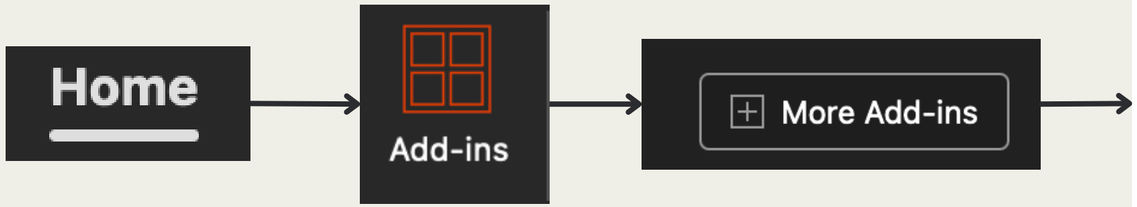
b. Interpret the slope coefficient attached to TD.

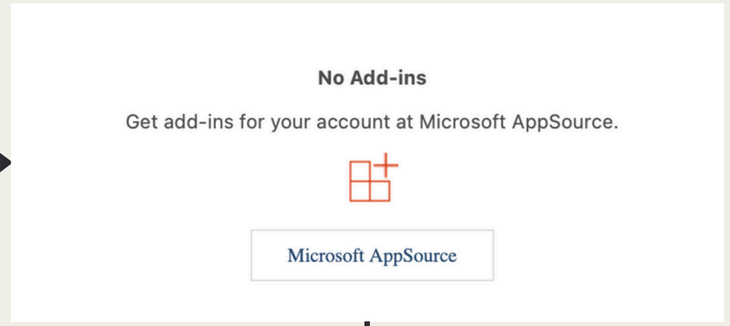
c. Player 8 earned 12.9895 million dollars during the season. According to the model, what is his predicted salary if PC = 70.6, TD = 34, and Age = 30?

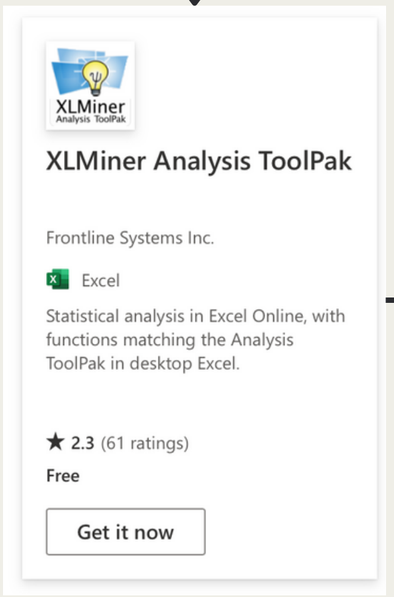
d. Player 16 earned 8.0073 million dollars during the season. According to the model, what is his predicted salary if PC = 65.5, TD = 28, and Age = 32?

e. Compute and interpret the residual salary for Player 8 and Player 16.

Q1A - DATA ANALYSIS TOOL ADD-IN

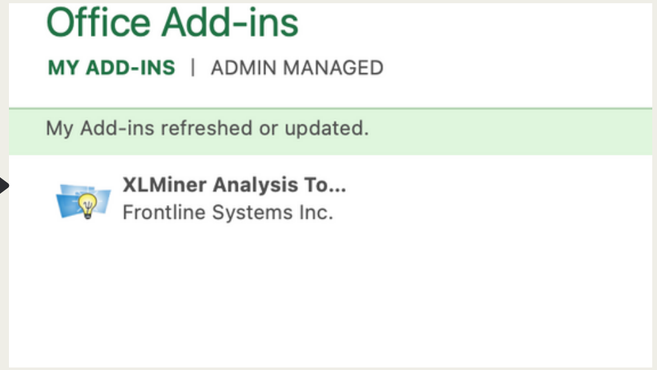






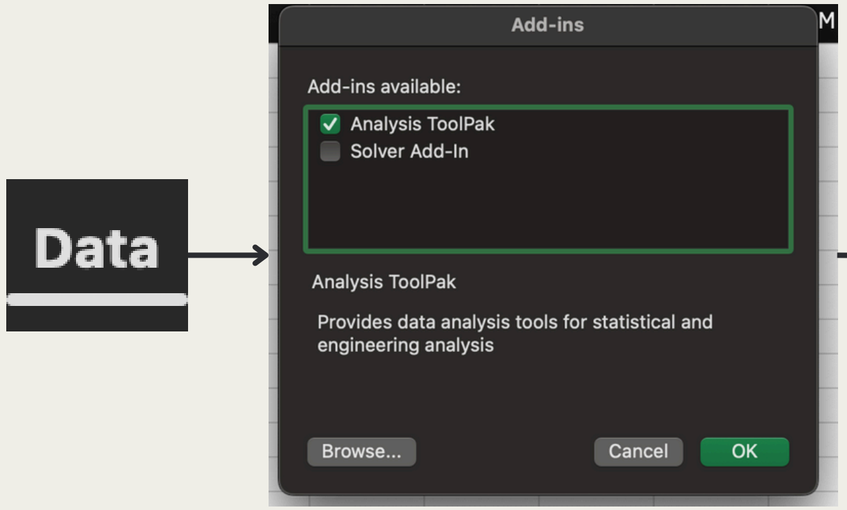
log in with uni email to get access to tool

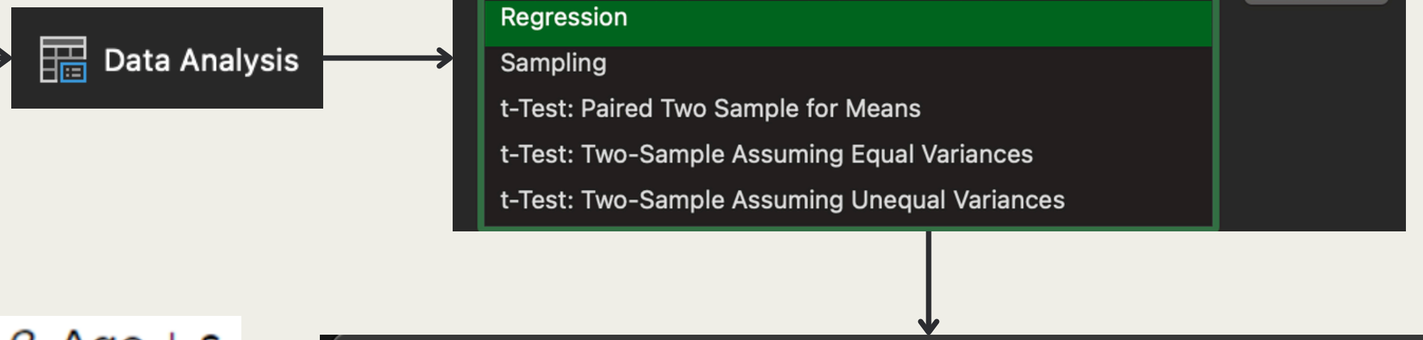




Refresh the Add-ins, it should show

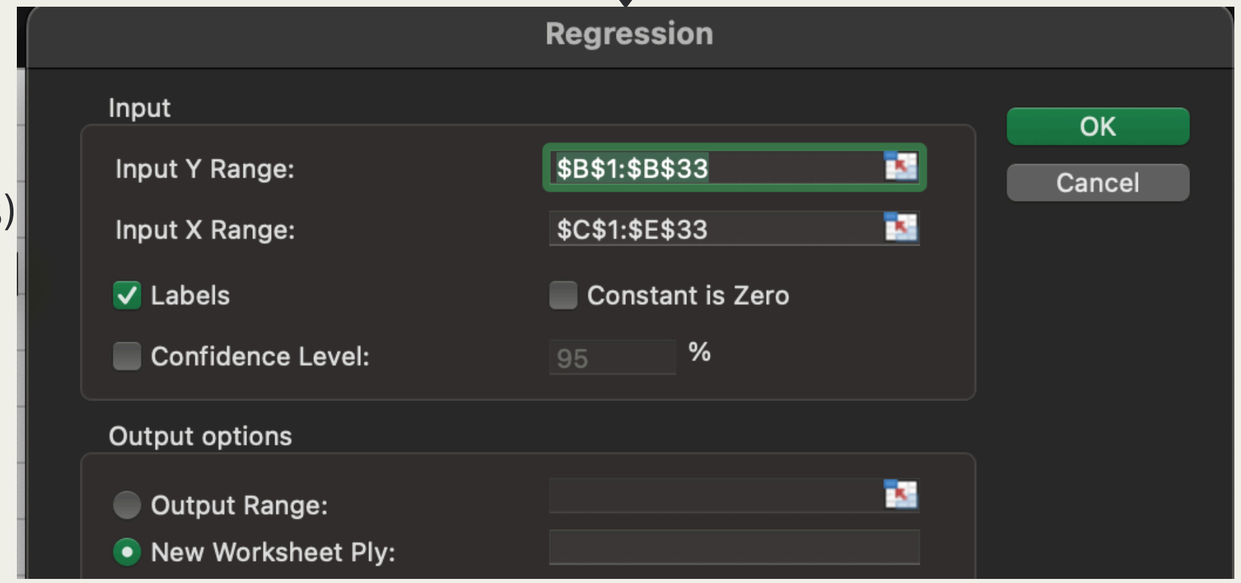
Q1A - REGRESSION DATA ANALYSIS





a. Estimate: Salary = $\beta_0 + \beta_1 PC + \beta_2 TD + \beta_3 Age + \varepsilon$.

1. **Y Range:** $B$1:$B$33 (include the “Salary” header)
2. **X Range:** $C$1:$E$33 (include the headers PC, TD, Age)
3. Check Labels ✓ (Excel knows the first row is headers)
4. Output the results to a new worksheet



Intercept (β_0)
PC (β_1)
TD (β_2)
Age (β_3)

Q 1 B

b. Interpret the slope coefficient attached to TD.

- In regression, each slope coefficient (β) tells us how much the Salary changes if that variable increases by 1 unit, while holding everything else constant
- The coefficient for TD (Touchdowns) in the solution is 0.7850

- So the interpretation is:
  - *“For every additional touchdown a quarterback throws, the model predicts his salary will increase by about 0.785 million dollars ($785,000), assuming pass completion rate and age stay the same.”*

Q1C, D, E

c. Player 8 earned 12.9895 million dollars during the season. According to the model, what is his predicted salary if PC = 70.6, TD = 34, and Age = 30?

d. Player 16 earned 8.0073 million dollars during the season. According to the model, what is his predicted salary if PC = 65.5, TD = 28, and Age = 32?

e. Compute and interpret the residual salary for Player 8 and Player 16.

Thank you!