UNIVERSIDAD DE LAS AMÉRICAS PUEBLA

P25 LEC3092-3 Econometrics 1

Dr. Miguel Reyes

1st Task. **Part 1**. 20250226

1. In a certain 2001 article, Freedman et al. present data on the relative velocity and distance of 24 galaxies, based on data from the Hubble Space Telescope. Following theoretical principles of astrophysics, it is possible to estimate the age of the universe using this data. To access the data in R: i) Install the 'gamair' package; ii) load the package into your session by typing 'library(gamair)'; iii) load the data using data(hubble). From here, you can access the data simply by typing 'hubble'.

   (a) Construct a scatter plot of the response variable (Y) 'velocity' versus the predictor variable (X) 'distance', and calculate the correlation coefficient between Y and X. Write your comments and observations on the graph and the correlation coefficient.

   (b) Perform a regression of Y versus X and obtain the relevant summary statistics, i.e., parameter estimates, standard errors, t-values, p-values, R^2, F-ratio, and the standard error of the regression. Comment on and interpret each of these.

   (c) Obtain 95% confidence intervals for β_0 and β_1, and interpret them.

   (d) Obtain intervals for the mean and prediction bands, both at 90%, for the estimated velocity when the distance is 5.5, 14.0, and 20.8. Provide an interpretation for each case.

   (e) Is there evidence to suggest that β_1 is different from zero? Is there evidence to suggest that β_0 is different from zero? Formally conduct a hypothesis test for each case and use the p-value to decide whether to reject the null hypothesis.

   (f) Construct and discuss the ANOVA table.

   (g) Use diagnostic plots to evaluate the validity of the model. Do you consider any transformation necessary for the predictor or the response?

2. Sometimes, it is necessary to fit a linear regression model where, from the outset (for theoretical or empirical reasons), it is known that the intercept is zero. This model is expressed as

$$y_i = \beta_1 x_i + e_i$$

   for $i = 1, 2, ..., n$.

   2.1 Show that in this context, the least squares estimator of β_1 is $\hat{\beta}_1 = \sum_i x_i y_i / \sum_i x_i^2$.

   2.2 Show that $\widehat{\beta_1}$ is an unbiased estimator for β_1 and that Var $\left[\widehat{\beta_1}\right] = \sigma^2 / \sum_i x_i^2$.

   2.3 Find an expression for $\widehat{\sigma^2}$. How many degrees of freedom are associated to it?

   2.4 Reconsider the 'hubble' dataset from Exercise 1. Let Y (velocity) be the response and X

(distance) be the predictor.

- - - 2.4.1 Perform a regression that passes through the origin. Find $\widehat{\beta_1}$ y $\widehat{\sigma^2}$. Find a 95% confidence interval and perform a test of hypothesis for $\beta_1$to verify whether there is evidence that $\beta_1 \neq 0$.

- - - 2.4.2 Plot the regression line on a scatter plot. Comment on the appropriateness of this model for the data and note any differences you find compared to the model fitted in Exercise 1. Use diagnostic plots to validate the model.

- - - 2.4.3 The regression model, $Y = \beta_1 X + \epsilon$, where Y is the velocity and X is the distance, is what astronomers essentially refer to as Hubble's Law. β_1 is the so-called Hubble constant, and β_1^{-1} provides an estimate of the age of the universe. The Hubble constant has units of km $\times$ second^{-1} $\times$ Mpc^{-1}. One megaparsec (Mpc) is 3.09×10^{19} km, so we need to divide the estimated value of β_1 by this amount to obtain the Hubble constant in units of second^{-1}. The approximate age of the universe in seconds will then be the inverse of this calculation. To obtain the age of the universe in years, execute the following in R:

R> Mpc <-3.09 * 10^{19}

R> ysec <- 60^2* 24 * 365.25

R> Mpcyear <- Mpc / ysec

R> 1 / ($\widehat{\beta_1}$ / Mpcyear).

What is the approximate age of the universe, expressed in years?

Specifications

- **Team Structure.** This task can be completed in teams of up to three members.
- **Report Requirements.** Write a report detailing your findings.
  - **General Expectations.** The report should provide a clear and structured explanation of how each subquestion was solved, rather than just presenting raw results or copied output from R. Include explanations, interpretations, and justifications where appropriate for each step taken.
  - **Cover Page.** The report must include a cover page or a header with the following information:
    - Course name
    - Instructor's name
    - Assignment title
    - Team members' full names
    - Submission date
  - **Formatting.** If using Word,

    * Paper size: US Letter;
    * Font: Aptos (Body), size 11;
    * Margins: 1 inch (top, bottom, right, left);
    * Line spacing: 1.5.
    * If using other editors, aim for similar formatting.
  - **Report Structure.** The report should be divided into two sections, one for each exercise. Each subquestion should be clearly labeled and answered with a logical explanation, not just pasted images or code output. Include essential elements such as graphs, tables, and statistical summaries where necessary. Use figure labels (e.g., "As shown in Figure 1, we can observe that...") and ensure they are properly referenced in the text.
  - **Writing Style**. Avoid unnecessary filler text or convoluted language. Be formal yet direct, concise, and to the point.

- **Submission Guidelines.** Each team member must upload the following to Blackboard: **1) The report in PDF format. 2) The R script containing the code solutions for the exercises.**

- **Deadline.** Friday, 07/03, 2025, 23:59.

- **Late Submissions.** Late submissions are accepted with a 10% penalty per day overdue. A submission is considered late starting on Saturday, 08/03/2025, at 00:00.