

# Estudio de caso # 5

Universidad Externado de Colombia

Departamento de Matemáticas

Estadística 2

Juan Sosa, Ph. D.

November 14, 2018

## Instrucciones generales

- Puede hacer el examen solo o puede asociarse con otra persona, entendiendo que la calificación del examen será la misma para ambas personas.
- El reporte final se debe enviar a más tardar el **viernes 23 de noviembre de 2018** a las **11:59 p.m.** a la cuenta de correo:  
`juan.sosa@uexternado.edu.co`.
- Reportar las cifras utilizando la **cantidad adecuada de decimales**, dependiendo de lo que se quiera mostrar y las necesidades del problema.
- Numerar figuras y tablas (<http://unilearning.uow.edu.au/report/1fi.html>) y proporcionarles un **tamaño adecuado** que no distorsione la información que estas contienen.
- El archivo del reporte final debe ser un **archivo pdf** con el siguiente formato: Letra Calibri, tamaño 12, interlineado sencillo con espacio entre párrafos y texto justificado. Márgenes: Normal. Tamaño: Carta. Orientación: Vertical.

- Especificar el software donde se llevó a cabo el computo e **incluir el código** correspondiente como un anexo al final del reporte con el siguiente formato: Letra Courier New, tamaño 10, interlineado sencillo.
- El objetivo principal de este trabajo es la claridad lógica y la interpretación de los resultados. **El informe no necesita ser extenso.** Recuerde ser minimalista escribiendo el reporte. Se deben incluir solo aquellos gráficos y tablas (¡y valores en la tabla!) que son relevantes para la discusión.
- Hacer el informe ya sea en inglés o español. No ambos!
- **Cualquier evidencia de plagio o copia se castigará severamente** tal y como el reglamento de la Universidad Externado de Colombia lo estipula.

Si está claro que (por ejemplo) dos estudiantes han trabajado juntos en una parte de un problema que vale 20 puntos, y cada respuesta habría ganado 16 puntos (si no hubiera surgido de una colaboración ilegal), entonces cada grupo recibirá 8 de los 16 puntos obtenidos colectivamente (para una puntuación total de 8 de 20), **y me reservo el derecho de imponer penalidades adicionales a mi discreción.**

Si un estudiante resuelve un problema por su cuenta y luego comparte su solución con cualquier otro estudiante (porque rutinariamente Usted hace esto, o por lástima, o bondad, o por cualquier motivo que pueda creer tener; no importa!), Usted es tan culpable de colaboración ilegal como la persona que tomó su solución, y ambos recibirán la misma penalidad. Este tipo de cosas es necesario hacerlas ya que muchas personas no hacen trampa, y debo asegurarme de que sus puntajes son obtenidos de manera genuina. En otras clases, personas perdieron la clase debido a una colaboración ilegal; **no deje que le suceda a Usted!**

## Pobreza Monetaria en Colombia 2016–2017

Una vez concluidas las dos fases del trabajo de la Misión para el Empalme de las Series de Empleo, Pobreza y Desigualdad (MESEP), el DANE asumió la responsabilidad de calcular y publicar los resultados de las dos mediciones oficiales de pobreza en Colombia: la Pobreza Monetaria y la Pobreza Multidimensional. El documento `MESEP.pdf` presenta la nueva metodología para la medición de pobreza monetaria en Colombia, que adopta cambios tanto en la línea de pobreza como en la construcción del agregado de ingreso del hogar. Toda la información se encuentra disponible en el siguiente enlace:

<https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/pobreza-y-desigualdad>

En el artículo *Aplicación de la regresión lineal en un problema de pobreza* (2013), los autores consideran un modelo de regresión lineal simple para caracterizar (explicar) la Pobreza Monetaria (PM) de 2011 en términos de la PM de 2010, en las trece principales ciudades de Colombia. En el presente caso de estudio, se replicará la misma caracterización para los años 2016–2017, pero esta vez para 23 departamentos y Bogotá D.C.; es decir, se quiere ajustar un modelo regresión lineal simple para evaluar el impacto de la PM de 2016 sobre la PM de 2017 usando la información de 23 departamentos y Bogotá D.C.. Para tal fin, considere las bases de datos `pobreza 2016.xlsx` y `pobreza 2017.xlsx` emitidas por el DANE.

1. Consolidar la información de la PM de 2016 y 2017 para los 23 departamentos y Bogotá D.C. ( $n = 24$  filas y dos columnas).
2. Clasificar las variables según su naturaleza y su escala de medición.
3. Identificar los cinco registros con mayores tasas de PM de 2016. Repetir para 2017. ¿Hay algún cambio?
4. Describir numérica y gráficamente cada una de las variables las variables por separado. Comentar los resultados obtenidos.
5. Hacer un dispersograma de la PM de 2017 (eje  $y$ ) frente a la PM de 2016 (eje  $x$ ). Calcular el coeficiente de correlación lineal de Pearson. Comentar los resultados obtenidos.

6. Considere el modelo de regresión lineal simple

$$\textbf{Modelo 1:} \quad y_i = \beta_1 + \beta_2 x_i + \epsilon_i, \quad \epsilon_i \stackrel{\text{iid}}{\sim} \text{N}(0, \sigma^2), \quad i = 1, \dots, n,$$

donde  $x_i$  y  $y_i$  son la PM de 2016 y 2017 del  $i$ -ésimo registro, respectivamente. Usando el método de máxima verosimilitud, demostrar paso a paso que:

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x} \quad \text{y} \quad \hat{\beta}_2 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

donde  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  y  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ .

**Nota:** Dado que  $\epsilon_i \stackrel{\text{iid}}{\sim} \text{N}(0, \sigma^2)$ , observe que cada  $y_i$  tiene distribución Normal con media  $\beta_1 + \beta_2 x_i$  y varianza  $\sigma^2$ , esto es,  $y_i \stackrel{\text{iid}}{\sim} \text{N}(\beta_1 + \beta_2 x_i, \sigma^2)$ . Por lo tanto la función de verosimilitud está dada por:

$$L(\beta_1, \beta_2, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - (\beta_1 + \beta_2 x_i))^2\right\}$$

7. Por medio de R (o cualquier otro software especializado de su preferencia), y usando una confiabilidad del 95%:

- (a) Calcular e interpretar  $\hat{\beta}_1$  y  $\hat{\beta}_2$ .
- (b) ¿El modelo en su totalidad resulta significativo? ¿Por qué? Interpretar los resultados obtenidos. Usar pruebas de hipótesis.
- (c) ¿El coeficiente  $\beta_1$  resulta significativo? ¿Por qué? Interpretar los resultados obtenidos. Usar tanto intervalos de confianza como pruebas de hipótesis.
- (d) ¿El coeficiente  $\beta_2$  resulta significativo? ¿Por qué? Interpretar los resultados obtenidos. Interpretar los resultados obtenidos. Usar tanto intervalos de confianza como pruebas de hipótesis.

8. Considere el modelo de regresión lineal simple (sin intercepto)

$$\textbf{Modelo 2:} \quad y_i = \beta_2 x_i + \epsilon_i, \quad \epsilon_i \stackrel{\text{iid}}{\sim} \text{N}(0, \sigma^2), \quad i = 1, \dots, n,$$

donde  $x_i$  y  $y_i$  son la PM de 2016 y 2017 del  $i$ -ésimo registro, respectivamente.

- (a) Completar la siguiente tabla:

Modelo	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}$	gl Reg	gl Err	SCR	SCE	$R^2$	$R^2_{adj}$
Modelo 1									
Modelo 2	NA								

- (b) Con base en la tabla anterior, ¿cuál modelo es mejor? ¿Por qué?
9. Calcular la desviación estandar (muestral) de la variable dependiente. Comparar este valor con la desviación estándar residual del modelo 2. ¿Hubo una decremento? Der ser así, ¿cuál fue el decremento en puntos porcentuales? ¿Qué dice esto acerca del modelo?
10. Evaluar los supuestos del modelo 2:
- Graficar los residuales estandarizados frente a los valores ajustados. ¿Este gráfico sugiere que la relación entre las variables es efectivamente lineal? ¿Por qué?
  - Graficar los residuales estandarizados frente a la variable independiente. ¿Este gráfico sugiere que la varianza del error es constante? ¿Por qué?
  - Graficar un diagrama cuantil-cuantil normal junto con las bandas de confianza para los residuales estandarizados. Probar el sistema de hipótesis correspondiente usando las pruebas de Shapiro-Wilk y de Kolmogorov-Smirnov. ¿Los residuales parecen provenir de una distribución Normal? ¿Por qué?
  - Probar si los residuales son independientes por medio de la prueba de Durbin-Watson. ¿El error parece satisfacer el supuesto de independencia? ¿Por qué?
  - ¿Se satisfacen todos los supuestos del modelo? ¿Por qué?
11. Graficar el modelo 2 sobre el dispersograma del numeral 5., y sobre este mismo gráfico, incluir los valores ajustados y los intervalos de confianza para varios valores del rango de la PM de 2016.
12. Usando el modelo 2, predecir el valor medio de la PM de 2017 para un departamento cuyo valor de PM de 2016 es 20%. Reportar el margen de error y el intervalo de confianza correspondiente. Interpretar los resultados obtenidos.
13. Usando el modelo 2, predecir el valor de la PM de 2017 para un departamento cuyo valor de PM de 2016 es 80%. Reportar el margen de error

y el intervalo de confianza correspondiente. Interpretar los resultados obtenidos.

14. Usando los residuales estandarizados y los residuales estudentizados del modelo 2, ¿existen evidencia de outliers? ¿Cuáles departamentos?
15. Usando los los DFFITS y la distancia de Cook del modelo 2, ¿existen evidencia de observaciones influyentes? ¿Cuáles departamentos?

## Measuring Stock Market Risk

One measure of the risk or volatility of an individual stock is the standard deviation of the total return (capital appreciation plus dividends) over several periods of time. Although the standard deviation is easy to compute, it does not take into account the extent to which the price of a given stock varies as a function of a standard market index, such as the S&P 500. As a result, many financial analysts prefer to use another measure of risk referred to as beta. Betas for individual stocks are determined by simple linear regression. The dependent variable is the total return for the stock and the independent variable is the total return for the stock market. Various sources use different approaches for computing betas. For instance, some sources subtract the return that could be obtained from a risk-free investment (e.g., T-bills) from the dependent variable and the independent variable before computing the estimated regression equation. Some also use different indexes for the total return of the stock market; for instance, Value Line computes betas using the New York Stock Exchange composite index.

For this case problem we will use the S&P 500 index<sup>1</sup> as the measure of the total return for the stock market, and an estimated regression equation will be developed using monthly data. The beta for the stock is the slope of the estimated regression equation ( $\beta_1$ ). The data contained in the file named **Beta.csv** provides the total return (capital appreciation plus dividends) over 36 months for eight widely traded common stocks and the S&P 500.

---

<sup>1</sup>The Standard & Poor's 500, often abbreviated as the S&P 500, or just the S&P, is an American stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ. The S&P 500 index components and their weightings are determined by S&P Dow Jones Indices.

The value of beta for the stock market will always be 1; thus, stocks that tend to rise and fall with the stock market will also have a beta close to 1. Betas greater than 1 indicate that the stock is more volatile than the market, and betas less than 1 indicate that the stock is less volatile than the market. For instance, if a stock has a beta of 1.4, it is 40% more volatile than the market, and if a stock has a beta of .4, it is 60% less volatile than the market.

You have been assigned to analyze the risk characteristics of these stocks (see the data set given in `Beta.csv`). Prepare a report that includes but is not limited to the following items.

1. Compute descriptive statistics for each stock and the S&P 500. Comment on your results. Which stocks are the most volatile?
2. Compute the value of beta for each stock. Which of these stocks would you expect to perform best in an up market? Which would you expect to hold their value best in a down market?
3. Comment on how much of the return for the individual stocks is explained by the market.