Práctica 3

Modelamiento estadístico

Desviación estándar

Es un estadístico fundamental para medir la dispersión de los datos, y en la regresión tiene un papel importante al momento de calcular la varianza asintótica de los estimadores y sus correspondientes errores estándares.

El programa STATA calcula la desviación estándar de los errores de la siguiente manera:

$$\hat{\sigma} = \sqrt{\frac{1}{N-K}\sum_{i=1}^{N}(\hat{e}_i)}$$

donde $\hat{e}_i = y_i - x_i'\hat{\beta}$. Este es un estimador consistente de la desviación estándar.

$$\hat{\sigma} \to^p \sigma$$

Para visualizar esta propiedad, vamos a crear un gran número de muestras para una variable aleatoria y_N, donde $y \sim N(10, 4)$ y estimaremos σ_y utilizando $\hat{\sigma}_y$ haciendo crecer N.

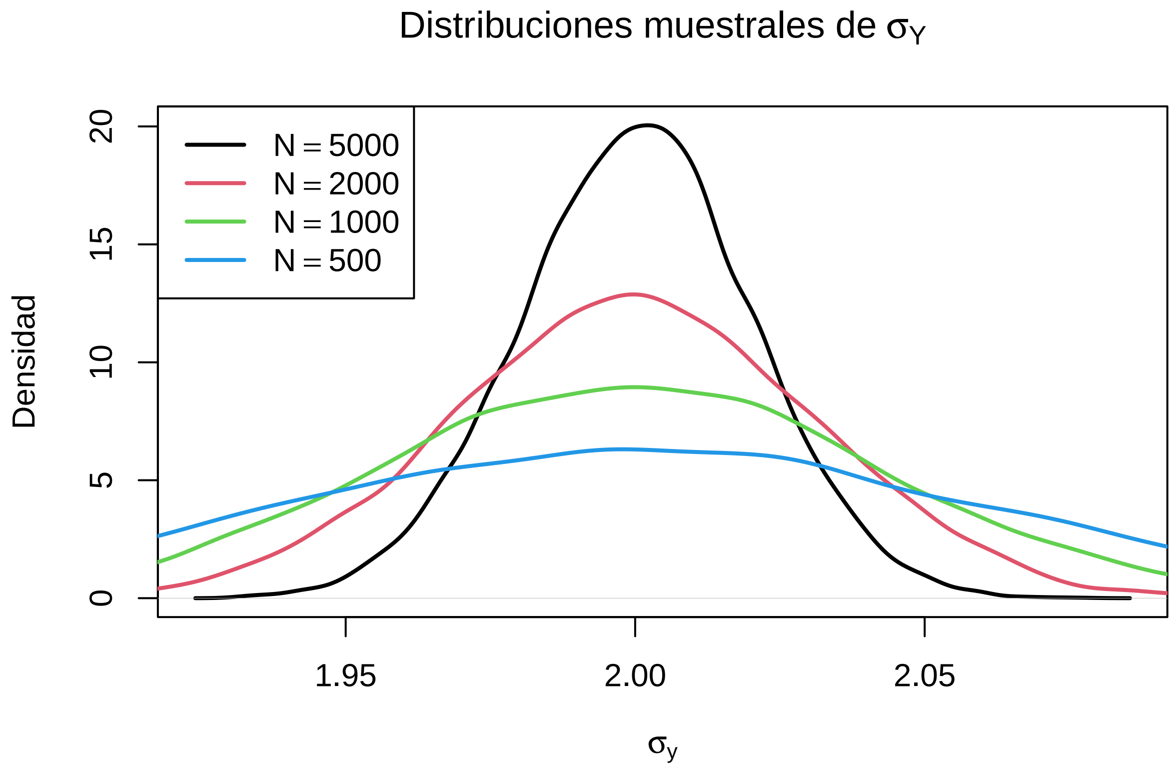
```r
# vector de tamaños de muestra
N <- c(5000, 2000, 1000, 500)

# graficar las desviaciones usando sd (standar deviations) para los diferentes tamaños de muestra
sq_y <- replicate(n = 10000, expr = sd(rnorm(N[1], 10, 2)))
plot(density(sq_y),
     main = expression("Distribuciones muestrales de" ~ sigma[Y]),
     xlab = expression(sigma[y]),
     ylab = expression("Densidad"),
     lwd = 2)

for (i in 2:length(N)) {
  sq_y <- replicate(n = 10000, expr = sd(rnorm(N[i], 10, 2)))
  lines(density(sq_y),
        col = i,
        lwd = 2)
}

# añadimos leyenda
legend("topleft",
       legend = c(expression(N == 5000),
                  expression(N == 2000),
                  expression(N == 1000),
                  expression(N == 500)),
```

```r
        col = 1:5,
        lwd = 2)
```



La distribución de $\hat{\sigma}$ se acerca al verdadero valor $\sigma = 2$ cuando N incrementa.

Ejercicio Práctico

Growth contiene data de crecimiento promedio para 65 países en el periodo 1960-1995, adicionalmente tiene variables potencialmente relacionadas con el crecimiento. Este data set proviene del artículo "Finance and the Sources of Growth" (JFE,2000) de Levine, Beck y Loayza.

Variable Definitions

Variable	Definition
Country_name	Name of country
growth	Average annual percentage growth of real Gross Domestic Product (GDP)* from 1960 to 1995.
rgdp60	The value of GDP* per capita in 1960, converted to 1960 US dollars
tradehare	The average share of trade in the economy from 1960 to 1995, measured as the sum of exports plus imports, divided by GDP; that is, the average value of $(X + M)/GDP$ from 1960 to 1995, where X = exports and M = imports (both X and M are positive).
yearsshcool	Average number of years of schooling of adult residents in that country in 1960
rev_coups	Average annual number of revolutions, insurrections (successful or not) and coup d'etats in that country from 1960 to 1995
assasinations	Average annual number of political assassinations in that country from 1960 to 1995 (per million population)
oil	= 1 if oil accounted for at least half of exports in 1960 = 0 otherwise

1. Visualice los datos y variables
2. Si estamos interesados en la relación entre `tradeshare` y `crecimiento`, realice un gráfico que ayude a visualizar dicha relación.
3. Calcule e interprete el coeficiente correlación de pearson :

$$\rho_{xy} = \frac{Cov(xy)}{\sigma_x \sigma_y}$$

Donde σ_x, σ_y son las deasviaciones estándar de cada variable.

4. Realice una estimación lineal del modelo:

$$growth_i = \beta_0 + \beta_1 tradeshare + u_i$$

5. Con la información previa, calcule manualmente el t estadístico, el valor crítico y el p-valor para un nivel de significancia de $\alpha = 0.05$.
6. Calcule e interprete el intervalo de confianza para el β estimado.

Un intervalo de confianza es un rango de valores, derivado de una muestra de datos, que con cierta probabilidad (nivel de confianza) contiene el verdadero valor de un parámetro poblacional.

En MCO ayuda a medir la incertidumbre alrededor de los coeficientes estimados.

Para un coeficiente estimado $\hat{\beta}$, su *intervalo de confianza* al $(1 - \alpha)\%$ se calcula como:

$$IC = \beta \pm c_\alpha * SE$$

7. Realice un gráfico ajustando una línea de regresión a los datos