ANEXO 3-A
LA LECTURA DE UNA ESPECIE DE COMPUTADORA

La presente sección busca echar un vistazo en el mundo del análisis de regresión. Presentamos, pues, tres extractos de salidas de computadora que provienen de los resultados de Lemelin y Polèse (1995). El programa que usamos es SAS. El primer extracto contiene los resultados relativos a la ecuación (6) de Lemelin y Polèse (1995):

$$\ln PURB_i = a + b \ln PTOT_i + c \ln GNPC_i + d\,(\ln GNPC_i) \qquad (6)$$

No se hace mención de los resultados del segundo extracto en Lemelin y Polèse (1995). Se trata del modelo truncado:

$$\ln PURB_i = a + b \ln PTOT_i + c \ln GNPC_i$$

Finalmente, el tercer extracto contiene los resultados relativos a la ecuación (1) de Lemelin y Polèse (1995):

$$\ln PLAR_i = \ln K + h \ln PURB_i \qquad (1)$$

Entre los elementos que podemos encontrar las salidas de computadora, mencionemos:

1. La variable dependiente de la regresión es *LPURB*, o sea ln *PURB*.
2. Este cuadro contiene los resultados de la estimación de los parámetros *a*, *b*, *c* y *d*.

3. Los parámetros se identifican por el nombre de la variable de la cual son coeficiente (en SAS, la variable INTERCEP, del inglés "intercept", refiera a la constante del modelo).
4. Los valores estimados de los parámetros y parámetros estándar *Beta*.

Los parámetros *Beta* son los valores que tendrían los coeficientes si se reemplazara las variables del modelo por variables estándar. Por ejemplo, se reemplazaría:

$$\ln PURB_i \text{ por } ZLPURB_i = \frac{\ln PURB_i - \overline{\ln PURB}}{s_{\ln PURB}},$$

$$\text{con } s_{\ln PURB} = \sqrt{\frac{1}{n-1}\sum_i \left(\ln PURB_i - \overline{\ln PURB}\right)^2}$$

$$\ln PTOT_i \text{ por } ZLPTOT_i = \frac{\ln PTOT_i - \overline{\ln PTOT}}{s_{\ln PTOT}},$$

$$\text{con } s_{\ln PTOT} = \sqrt{\frac{1}{n-1}\sum_i \left(\ln PTOT_i - \overline{\ln PTOT}\right)^2}$$

etc.

y el parámetro *Beta* relacionado a ln *PTOT* se da con

$$Beta_{\ln PTOT} = \hat{b}\frac{s_{\ln PTOT}}{s_{\ln PURB}}$$

5. Error tipo (desviación estándar de muestro) de los valores estimados de los parámetros: $s_{\hat{a}}$, $s_{\hat{b}}$, $s_{\hat{c}}$ y $s_{\hat{d}}$.
6. Valor del *t* de Student para la hipótesis que el valor del parámetro sea cero; tenemos $t = \frac{\hat{a}}{s_{\hat{a}}}$, etc.
7. Probabilidad crítica correspondiente al valor del *t*: es el umbral de significanza que se tendría que tomar para que el valor crítico correspondiente sea igual al valor

calculado del t; es equivalente al resultado de la función `TDIST` en Excel.

8. El cuadro *Analysis of variance* (análisis de varianza) detalle el cálculo del R^2.
9. Valores de las sumas de cuadrados que entran en el cálculo del R^2 y sirven para construir tests F de Fisher:
   - Línea "Model": SSM
     = variabilidad de la cual el modelo de cuenta
   - Línea "Error": SSR = variabilidad residual
   - Línea "C Total": SST = variabilidad total
10. Número de grados de libertad (DF, "degrees of freedom") asociado a cada suma de cuadrados:
    - SSM: $k-1$
    - SSR: $n-k$
    - SST: $n-1$
11. Desviaciones cuadráticas promedias ("Mean squares"); se cálculan con división entre el número de grados de libertad correspondiente:
    - Modelo: $\dfrac{SSM}{k-1}$
    - Residuos: $MSE = \dfrac{SSR}{n-k}$; es el valor estimado de la varianza del término aleatorio, σ^2.
12. Valor del F de Fisher para la hipótesis nula que todos los parámetros sean cero, excepto la constante. Se cálcula con

$$F(k-1; n-k) = \frac{SSM / (k-1)}{SSR / (n-k)}$$

13. Probabilidad crítica correspondiente al valor calculado del F.
14. Coeficiente de determinación multiple:

$$R^2 = \frac{SSM}{SST} = 1 - \frac{SSR}{SST}$$

15. Coeficiente de determinación multiple ajustado:

$$\overline{R}^2 = 1 - \frac{SSR/(n-k)}{SST/(n-1)}$$

16. Evaluación de la capacidad de predicción del modelo:
   - Raíz cuadrada del error cuadrático promedio
     “Root MSE” $= \sqrt{MSE}$
   - Valor promedio de la variable dependiente: m_y.
   - “Coeficiente de variación” (C.V.); en realidad, se trata del coeficiente de variación en porcentage:

$$= 100 \frac{\sqrt{MSE}}{m_y}$$

17. Test de la hipótesis que el coeficiente de ln *PTOT* sea igual a 1. Este test se hace con el *F* de Fisher. Puesto que el test *F* y el test *t* de Student se basan en el mismo modelo muestral (modelo clásico de la regresión lineal normal), el resultado es absolutamente idéntico: la probabilidad crítica es 0.3144.

Especie del logicial SAS

```
                         PRIMATE moins HK ; UN POP DIV et GNPC                15
                                            18:19 Tuesday, August 29, 1995

Model: EQPURB
Dependent Variable: LPURB

                           Analysis of Variance

                               Sum of         Mean
        Source          DF    Squares       Square      F Value       Prob>F

        Model            3   88.81851     29.60617      409.878       0.0001
        Error           60    4.33390      0.07223
        C Total         63   93.15241

            Root MSE       0.26876     R-square       0.9535
            Dep Mean       9.47138     Adj R-sq       0.9511
            C.V.           2.83760

                           Parameter Estimates

                      Parameter      Standard    T for H0:
      Variable  DF     Estimate         Error   Parameter=0    Prob > |T|

      INTERCEP   1    -4.602824    0.87667347        -5.250        0.0001
      LPTOT      1     0.971663    0.02793210        34.787        0.0001
      LGNPC      1     0.915958    0.20762625         4.412        0.0001
      SQLGNPC    1    -0.045267    0.01344728        -3.366        0.0013
________________________________________________________________________________
                         PRIMATE moins HK ; UN POP DIV et GNPC                16
                                            18:19 Tuesday, August 29, 1995

Dependent Variable: LPURB
Test: PTOTEG1   Numerator:      0.0743  DF:    1   F value:   1.0292
                Denominator:  0.072232  DF:   60   Prob>F:    0.3144
```

Digresión: el aspecto de la relación entre la población urbana y el PIB per cápita

La ecuación (6) es una relación cuadrática entre los logaritmos de las variables. Sin embargo, ¿cuál es el aspecto real de la relación entre la población urbana y el PIB per cápita?

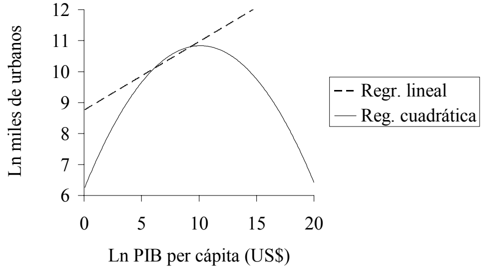
Antes de la era de las hojas de cálculo electrónicas (como Excel), cuando se deseaba conocer el aspecto de una función sin tener que calcular un gran número de valores, se recurría al cálculo diferencial y álgebra. Era necesario conocer las derivadas primeras y segundas de la función, luego resolver las ecuaciones que permitían determinar los puntos de intersección con el eje de las abscisas, sus máximas y mínimas, sus puntos de inflexión, para finalmente determinar el signo de la función y de sus derivadas entre estos puntos de referencia. Hoy en día, sólo basta simular en una hoja de cálculo.

Por lo que trata de la ecuación (6), los resultados presentados más arriba nos muestran que el coeficiente *b* es aproximadamente igual a 1. Esto implica que, todo siendo igual de un lado (particularmente el PIB per cápita), la población urbana es más o menos proporcional a la población total. Por lo tanto, nos concentraremos en el aspecto de la relación entre el PIB per cápita y la población urbana, dada una población total. Fijamos la cifra de la población en su valor promedio en la muestra (67.6 millones de habitantes) y hicimos variar el PIB per cápita entre cero y el valor inverosímil de US$ 100.000. Para cada valor, calculamos cuál fue la población urbana predicha por el modelo. Los resultados se reproducen a continuación en graficos.

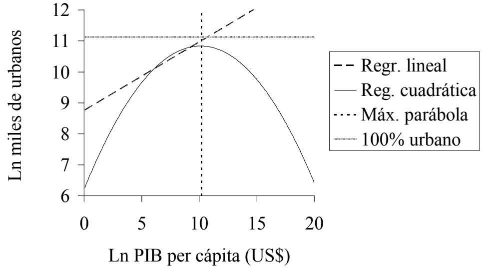
Los puntos que queremos principalmente destacar son:

- El modelo que representa la ecuación (6) es un modelo descriptivo el cual es válido solo al interior de un cierto campo de variación.
- El aspecto visual de la curva depende de la selección de las escalas (natural o logarítmica).

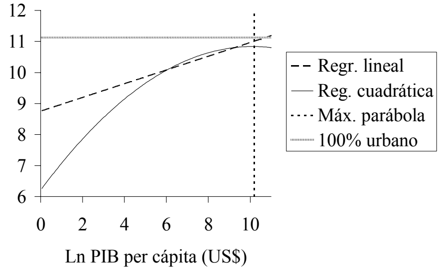
Población urbana calculada[177]



Población urbana calculada

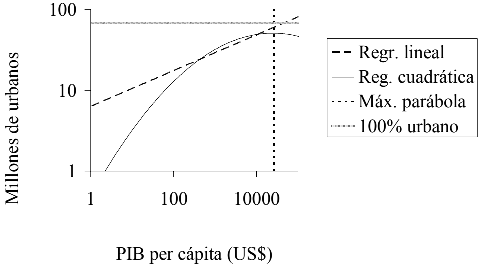


Población urbana calculada

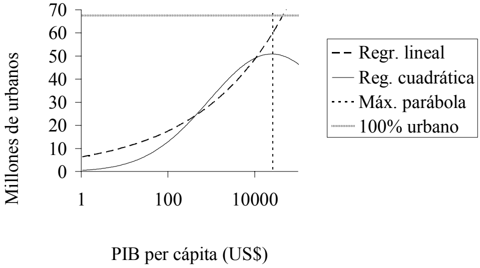


[177] Cálculo hecho con una población total de 67.6 millones.

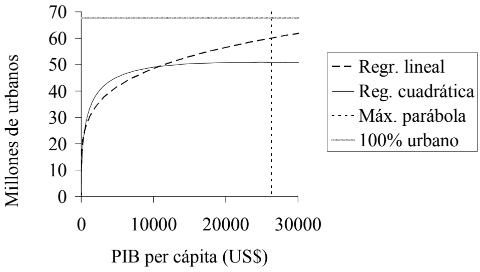
Población urbana calculada (escalas logarítmicas)



Población urbana calculada (escala horizontal logarítmica)



Población urbana calculada



Tasas de urbanización observadas y calculadas