

## Tarea 2

**Entrega: 9 de Octubre**

**Deberán mandar por email el do-file que genera sus respuestas**

Para esta tarea utilizaremos la base de datos *nba\_salaries\_12.dta* donde tenemos información de los jugadores de la NBA. Los datos que se recopilan son: su salario (*wage*), edad (*age*), estado marital (*marr*), años de experiencia (*exper*), promedio de juegos por año (*games*), promedio de minutos jugados por partido (*avgmin*), dummies para indicar la posición en la que juegan {*Center*, *Guard*, *Forward* }, puntos por juego (*points*), rebotes por juego (*rebounds*), asistencias por juego (*assists*), entre otras variables. Una descripción más detallada de esta base de datos la pueden encontrar en el archivo *nbasal\_desc* que pueden abrir con cualquier editor de texto (e.g. WordPad). Con estos datos buscaremos construir estimaciones que describan qué variables influyen en los niveles de salario de los jugadores.

1. Crea una tabla de estadísticos descriptivos que incluya número de observaciones, media, error estándar, valor mínimo y máximo de las siguientes variables: *age*, *wage*, *exper*, *avgmin*, *black*, *guard*, *forward*, *center*, *points*, *rebounds*, *assists* [*Stata Tip*: Utiliza el comando summarize]
2. Ahora realiza las estimaciones correspondientes y completa la *Tabla 1* que está al final de la tarea. Recuerda que debes reportar los coeficientes y los errores estandar. [*Stata Tip*: En algunos casos necesitarás generar nuevas variables. Para esto necesitarás utilizar el comando gen.]
3. Explica ¿por qué cambió tanto el coeficiente de *age* entre la primera y la segunda estimación? Deberás justificar claramente tu respuesta.
4. Utilizando los resultados de tus estimaciones da una interpretación lo más específica posible para los siguientes coeficientes:
  - a) *Age* en las columnas (3)

- b) *Points* en la columna (3)
  - c) *Black* en la columna (3)
  - d)  $\ln(\text{avgmin})$  en la columna (3)
  - e) *Exper* en la columna (4)
  - f) *Forward* en la columna (4)
  - g) *Rebounds* en la columna (4)
  - h)  $\ln(\text{avgmin})$  en la columna (4)
5. ¿Qué pasaría con la estimación (4) si trataras de agregar *Guard* como control? ¿Por qué cree que ocurre esto?
6. Haz una gráfica de dispersión de  $\ln(\text{Wage})$  vs. *Points*. De acuerdo a tu gráfica, ¿qué tipo de errores tiene más sentido usar? Justifica tu respuesta. [*Stata Tip*: Deberás usar el comando **twoway (scatter x y)**]
7. Modifica la estimación (4) agregando como controles  $\text{Points}^2$  y  $\text{Points}^3$ . Reporta tus resultados en formato de ecuación. Utilizando este resultado evalúa si la relación entre puntos por partido y salario es lineal.
8. Michael Jordan Jr. se quiere comprar una casa nueva. Michael es un jugador afroamericano con 23 años de edad, 3 años de experiencia, juega en la posición de *Guard*, que en promedio por partido anota 20 puntos, gana 8 rebotes, da 5 asistencias y juega 30 minutos.
- Él prevé que dentro de 3 años va a aumentar sus números en 4 puntos, 1 rebote y 1 asistencia por partido, jugando la misma cantidad de minutos. Para comprar su casa nueva necesita aumentar su salario, por lo menos, 25 %. Utilizando los datos de la estimación de la columna (4), ¿Le recomendarías a Michael que se comprara la casa o no?
- Para hacerle una recomendación a Michael deberás ser muy profesional, utilizar tus datos, plantear una prueba de hipótesis y reportar tus resultados que soporte tu recomendación con un lenguaje estadístico correcto (para que tu profesor te evalúe). Pero además deberás reportar tu análisis y resultados en un lenguaje que sea comprensible

para alguien como Michael que no es experto en estadística [*Stata tip*: necesitarás el comando `lincom`].

9. Un comentarista deportivo dice durante el noticiero que los *Guards* afroamericanos sufren de racismo y son discriminados en términos salariales en la liga por debajo de otros jugadores con características similares. Utilizando la regresión de la columna (4), o una versión modificada de esta si es necesario, da evidencia para respaldar o contradecir si:
  - a) Los *Guards* afroamericanos ganan menor salario que los *Guards* no afroamericanos
  - b) Los *Guards* afroamericanos ganan menor salario que cualquier otro jugador (que no sea *Guard* afroamericano)

En cada caso deberás plantear las pruebas de hipótesis relevante y evaluarlas. (Nota: si estimas otra regresión reportala en formato de ecuación)

10. Un consultor de deportes te asegura que en la regresión de la columna (4) no existe ninguna variable omitida que pueda estar correlacionada con *avgmin*. Supón que le crees, ¿qué otro argumento podrías utilizar para indicar que podría existir sesgo en la estimación de ese coeficiente?
11. La estimación de la columna (4) parece indicar que el número de rebotes no es relevante para el salario. Sin embargo, esta variable puede ser sensible a la posición del jugador. Modifica la estimación de la columna (4) para evaluar si el efecto del aumento de un rebote en el salario es igual para la posición de *Centro* (que típicamente corresponde a los jugadores más altos que se dedican a ganar rebotes) que para la posición *Forward*. Indica claramente qué prueba de hipótesis estás evaluando y por qué. Reporta los resultados de tu nueva estimación en la columna (5). Si agregaste o quitaste alguna variable deberás incluirlo en el renglón adicional que hay en la tabla. Si necesitas agregar más variables puedes agregar renglones.

Tabla 1

	(1)	(2)	(3)	(4)
Variable dependiente:	<i>Wage</i>	<i>Wage</i>	<i>Wage</i>	$\ln(Wage)$
<i>Age</i>				
<i>Exper</i>	—			
<i>Points</i>	—	—		
<i>Rebounds</i>	—	—		
<i>Assists</i>	—	—		
<i>Black</i>	—	—		
$\ln(avgmin)$	—	—		
<i>Center</i>	—	—		
<i>Forward</i>	—	—		
	—	—	—	—
Constante				
<i>N</i>				
<i>R</i> <sup>2</sup>				

\*\*\*Significativo al 1 %, \*\*Significativo al 5 %, \*Significativo al 10 %.

Se asumen errores heteroscedásticos en las estimaciones.