

MODELOS LINEALES

SEGUNDO TALLER

Fernando Massa; Bruno Bellagamba

Jueves 18 de abril 2024



FACULTAD DE
CIENCIAS ECONÓMICAS
Y DE ADMINISTRACIÓN

IESTA INSTITUTO
DE ESTADÍSTICA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

ANTES DE ARRANCAR...

- 1 Cambie el directorio de trabajo a la carpeta donde tiene los datos *condados.xlsx*.

ANTES DE ARRANCAR...

- ➊ Cambie el directorio de trabajo a la carpeta donde tiene los datos *condados.xlsx*.
- ➋ Abra un nuevo script (o .Rmd) guardelo con el nombre *segundo_taller.r*.

ANTES DE ARRANCAR...

- ❶ Cambie el directorio de trabajo a la carpeta donde tiene los datos *condados.xlsx*.
- ❷ Abra un nuevo script (o .Rmd) guardelo con el nombre *segundo_taller.r*.
- ❸ Cargue el conjunto de datos *condados.xlsx*

ANTES DE ARRANCAR...

- ❶ Cambie el directorio de trabajo a la carpeta donde tiene los datos *condados.xlsx*.
- ❷ Abra un nuevo script (o .Rmd) guardelo con el nombre *segundo_taller.r*.
- ❸ Cargue el conjunto de datos *condados.xlsx*

ANTES DE ARRANCAR...

- 1 Cambie el directorio de trabajo a la carpeta donde tiene los datos *condados.xlsx*.
- 2 Abra un nuevo script (o .Rmd) guardelo con el nombre *segundo_taller.r*.
- 3 Cargue el conjunto de datos *condados.xlsx*

Seleccionemos las observaciones del estado de **Alabama**.

ANTES DE ARRANCAR...

- ❶ Cambie el directorio de trabajo a la carpeta donde tiene los datos *condados.xlsx*.
- ❷ Abra un nuevo script (o .Rmd) guardelo con el nombre *segundo_taller.r*.
- ❸ Cargue el conjunto de datos *condados.xlsx*

Seleccionemos las observaciones del estado de **Alabama**.

La idea de hoy consiste en ajustar un modelo lineal múltiple para explicar los valores de la esperanza de vida (*esp_vida*) en los condados del estado que elegimos en base a:

ANTES DE ARRANCAR...

- 1 Cambie el directorio de trabajo a la carpeta donde tiene los datos *condados.xlsx*.
- 2 Abra un nuevo script (o .Rmd) guardelo con el nombre *segundo_taller.r*.
- 3 Cargue el conjunto de datos *condados.xlsx*

Seleccionemos las observaciones del estado de **Alabama**.

La idea de hoy consiste en ajustar un modelo lineal múltiple para explicar los valores de la esperanza de vida (*esp_vida*) en los condados del estado que elegimos en base a:

- el porcentaje de adultos con diabetes (*pct_diabetes*)
- el porcentaje de adultos con obesidad (*pct_obesidad*)
- el porcentaje de personas con estudios terciarios (*pct_est_terc*)
- el ingreso mediano de los hogares (*ing_med_hog*)
- el porcentaje de nacimientos con bajo peso (*pct_bpeso*)

Queremos ajustar un modelo que explique la esperanza de vida según el ingreso mediano de los hogares incluyendo también las demás variables explicativas.

Queremos ajustar un modelo que explique la esperanza de vida según el ingreso mediano de los hogares incluyendo también las demás variables explicativas.

- 1 Ajuste el modelo de regresión lineal múltiple.

Queremos ajustar un modelo que explique la esperanza de vida según el ingreso mediano de los hogares incluyendo también las demás variables explicativas.

- 1 Ajuste el modelo de regresión lineal múltiple.
- 2 ¿Cuánto vale ahora la estimación del coeficiente asociado al ingreso?

Queremos ajustar un modelo que explique la esperanza de vida según el ingreso mediano de los hogares incluyendo también las demás variables explicativas.

- 1 Ajuste el modelo de regresión lineal múltiple.
- 2 ¿Cuánto vale ahora la estimación del coeficiente asociado al ingreso?
- 3 Interpretelo.

Queremos ajustar un modelo que explique la esperanza de vida según el ingreso mediano de los hogares incluyendo también las demás variables explicativas.

- 1 Ajuste el modelo de regresión lineal múltiple.
- 2 ¿Cuánto vale ahora la estimación del coeficiente asociado al ingreso?
- 3 Interpretelo.
- 4 Interprete las estimaciones de los demás coeficientes.

Queremos ajustar un modelo que explique la esperanza de vida según el ingreso mediano de los hogares incluyendo también las demás variables explicativas.

- 1 Ajuste el modelo de regresión lineal múltiple.
- 2 ¿Cuánto vale ahora la estimación del coeficiente asociado al ingreso?
- 3 Interpretelo.
- 4 Interprete las estimaciones de los demás coeficientes.
- 5 ¿Cuánto vale el R^2 ?

Queremos ajustar un modelo que explique la esperanza de vida según el ingreso mediano de los hogares incluyendo también las demás variables explicativas.

- 1 Ajuste el modelo de regresión lineal múltiple.
- 2 ¿Cuánto vale ahora la estimación del coeficiente asociado al ingreso?
- 3 Interpretelo.
- 4 Interprete las estimaciones de los demás coeficientes.
- 5 ¿Cuánto vale el R^2 ?
- 6 Realice el gráfico de \hat{Y} contra $\hat{\varepsilon}$. ¿Qué observa?

Queremos ajustar un modelo que explique la esperanza de vida según el ingreso mediano de los hogares incluyendo también las demás variables explicativas.

- 1 Ajuste el modelo de regresión lineal múltiple.
- 2 ¿Cuánto vale ahora la estimación del coeficiente asociado al ingreso?
- 3 Interpretelo.
- 4 Interprete las estimaciones de los demás coeficientes.
- 5 ¿Cuánto vale el R^2 ?
- 6 Realice el gráfico de \hat{Y} contra $\hat{\varepsilon}$. ¿Qué observa?

A partir del modelo estimado en la etapa anterior:

- ¿Es el modelo significativo?

A partir del modelo estimado en la etapa anterior:

- ¿Es el modelo significativo? ¿A qué nivel de significación?

A partir del modelo estimado en la etapa anterior:

- ¿Es el modelo significativo? ¿A qué nivel de significación?
- ¿Qué variables explicativas son significativas al 5%?

A partir del modelo estimado en la etapa anterior:

- ¿Es el modelo significativo? ¿A qué nivel de significación?
- ¿Qué variables explicativas son significativas al 5%? ¿Y al 1%?

A partir del modelo estimado en la etapa anterior:

- ¿Es el modelo significativo? ¿A qué nivel de significación?
- ¿Qué variables explicativas son significativas al 5%? ¿Y al 1%?
- ¿Con cuántos grados de libertad se están realizando estas pruebas estas pruebas?

A partir del modelo estimado en la etapa anterior:

- ¿Es el modelo significativo? ¿A qué nivel de significación?
- ¿Qué variables explicativas son significativas al 5%? ¿Y al 1%?
- ¿Con cuántos grados de libertad se están realizando estas pruebas estas pruebas?
- Investigue qué hace la función *drop1* y utilícela para obtener los estadísticos F de significación individual de cada variable.

A partir del modelo estimado en la etapa anterior:

- ¿Es el modelo significativo? ¿A qué nivel de significación?
- ¿Qué variables explicativas son significativas al 5%? ¿Y al 1%?
- ¿Con cuántos grados de libertad se están realizando estas pruebas estas pruebas?
- Investigue qué hace la función *drop1* y utilícela para obtener los estadísticos *F* de significación individual de cada variable.
- ¿Es posible afirmar que los coeficientes de las variables *pct_diabetes*, *pct_bpeso* y *pct_est_ter* son simultáneamente cero?

A partir del modelo estimado en la etapa anterior:

- ¿Es el modelo significativo? ¿A qué nivel de significación?
- ¿Qué variables explicativas son significativas al 5%? ¿Y al 1%?
- ¿Con cuántos grados de libertad se están realizando estas pruebas estas pruebas?
- Investigue qué hace la función *drop1* y utilícela para obtener los estadísticos *F* de significación individual de cada variable.
- ¿Es posible afirmar que los coeficientes de las variables *pct_diabetes*, *pct_bpeso* y *pct_est_ter* son simultáneamente cero?
- ¿Es posible decir que por cada unidad que aumenta el porcentaje de personas con obesidad (dejando todo lo demás constante), la esperanza de vida del condado desciende en promedio 1 meses?

A partir del modelo estimado en la etapa anterior:

- ¿Es el modelo significativo? ¿A qué nivel de significación?
- ¿Qué variables explicativas son significativas al 5%? ¿Y al 1%?
- ¿Con cuántos grados de libertad se están realizando estas pruebas estas pruebas?
- Investigue qué hace la función *drop1* y utilícela para obtener los estadísticos *F* de significación individual de cada variable.
- ¿Es posible afirmar que los coeficientes de las variables *pct_diabetes*, *pct_bpeso* y *pct_est_ter* son simultáneamente cero?
- ¿Es posible decir que por cada unidad que aumenta el porcentaje de personas con obesidad (dejando todo lo demás constante), la esperanza de vida del condado descende en promedio 1 meses?
- ¿Qué le dice el intervalo de confianza el 95% del coeficiente asociado a la variable *pct_obesidad*?

- Realice esta prueba de hipótesis:

$$H_0) \quad \beta_{pct_est_ter} \leq 0$$

$$H_1) \quad \beta_{pct_est_ter} > 0$$

¿Qué conclusión obtiene?

- Realice esta prueba de hipótesis:

$$H_0) \quad \beta_{pct_est_ter} \leq 0$$

$$H_1) \quad \beta_{pct_est_ter} > 0$$

¿Qué conclusión obtiene?

- Empleando alguna variante de bootstrap, cree un intervalo de confianza al 95% para el R^2

- Realice esta prueba de hipótesis:

$$H_0) \quad \beta_{pct_est_ter} \leq 0$$

$$H_1) \quad \beta_{pct_est_ter} > 0$$

¿Qué conclusión obtiene?

- Empleando alguna variante de bootstrap, cree un intervalo de confianza al 95% para el R^2
- Realice la predicción de la esperanza de vida para un condado con 35% de personas obesas y con un ingreso mediano de 50000.

- Realice esta prueba de hipótesis:

$$H_0) \quad \beta_{pct_est_ter} \leq 0$$

$$H_1) \quad \beta_{pct_est_ter} > 0$$

¿Qué conclusión obtiene?

- Empleando alguna variante de bootstrap, cree un intervalo de confianza al 95% para el R^2
- Realice la predicción de la esperanza de vida para un condado con 35% de personas obesas y con un ingreso mediano de 50000.
- Realice un intervalo de confianza para la predicción del punto anterior.

- Realice esta prueba de hipótesis:

$$H_0) \quad \beta_{pct_est_ter} \leq 0$$

$$H_1) \quad \beta_{pct_est_ter} > 0$$

¿Qué conclusión obtiene?

- Empleando alguna variante de bootstrap, cree un intervalo de confianza al 95% para el R^2
- Realice la predicción de la esperanza de vida para un condado con 35% de personas obesas y con un ingreso mediano de 50000.
- Realice un intervalo de confianza para la predicción del punto anterior.
- Realice la predicción de la esperanza de vida para un **grupo de condados** con 35% de personas obesas y con un ingreso mediano de 50000.

- Realice esta prueba de hipótesis:

$$H_0) \quad \beta_{pct_est_ter} \leq 0$$

$$H_1) \quad \beta_{pct_est_ter} > 0$$

¿Qué conclusión obtiene?

- Empleando alguna variante de bootstrap, cree un intervalo de confianza al 95% para el R^2
- Realice la predicción de la esperanza de vida para un condado con 35% de personas obesas y con un ingreso mediano de 50000.
- Realice un intervalo de confianza para la predicción del punto anterior.
- Realice la predicción de la esperanza de vida para un **grupo de condados** con 35% de personas obesas y con un ingreso mediano de 50000.
- Realice un intervalo de confianza para la predicción del punto anterior.

- Realice esta prueba de hipótesis:

$$H_0) \quad \beta_{pct_est_ter} \leq 0$$

$$H_1) \quad \beta_{pct_est_ter} > 0$$

¿Qué conclusión obtiene?

- Empleando alguna variante de bootstrap, cree un intervalo de confianza al 95% para el R^2
- Realice la predicción de la esperanza de vida para un condado con 35% de personas obesas y con un ingreso mediano de 50000.
- Realice un intervalo de confianza para la predicción del punto anterior.
- Realice la predicción de la esperanza de vida para un **grupo de condados** con 35% de personas obesas y con un ingreso mediano de 50000.
- Realice un intervalo de confianza para la predicción del punto anterior.
- ¿Qué diferencias hay entre estos intervalos?

- Realice esta prueba de hipótesis:

$$H_0) \quad \beta_{pct_est_ter} \leq 0$$

$$H_1) \quad \beta_{pct_est_ter} > 0$$

¿Qué conclusión obtiene?

- Empleando alguna variante de bootstrap, cree un intervalo de confianza al 95% para el R^2
- Realice la predicción de la esperanza de vida para un condado con 35% de personas obesas y con un ingreso mediano de 50000.
- Realice un intervalo de confianza para la predicción del punto anterior.
- Realice la predicción de la esperanza de vida para un **grupo de condados** con 35% de personas obesas y con un ingreso mediano de 50000.
- Realice un intervalo de confianza para la predicción del punto anterior.
- ¿Qué diferencias hay entre estos intervalos?

¿Preguntas?

Muchas Gracias