

Prediciendo Alturas

ExactasPrograma - Datos

Facultad de Ciencias Exactas y Naturales, UBA

Invierno 2020

¿Cuánto medirá al ser adult^e?

¿Qué?

- Una excusa para hablar de datos y programar.
- Ejemplo elegido por razones históricas: *Regression Towards Mediocrity in Hereditary Stature*. (Galton, 1886)
- Datos y predicción.
- Conocimiento general: ideas simples pero poderosas
- Pensar para qué puede servirme.

Vamos de nuevo: la dulce espera



¿Cuánto medirá al ser adulte?

Es un varón



¿Cuánto medirá al ser adulte?

Es un varón y la mamá es bajita



Es un varón



La mamá es bajita

¿Cuánto medirá al ser adulta?

LA pregunta

¿Cuánto medirá al ser adult^e?

Misma pregunta con información adicional

La pregunta: ¿Cuánto medirá al ser adult^e?

Información Adicional: Será varón

Misma pregunta con más información adicional

La pregunta: ¿Cuánto medirá al ser adult^e?

Información Adicional: Será varón y la mamá es bajita

Misma pregunta con otra información (adicional ?)

La pregunta: ¿Cuánto medirá al ser adult^e?

Información Adicional: Será varón y la mamá mide 156.5 cm.

Pasito a Pasito (suave, suavequito)

La pregunta: ¿Cuánto medirá al ser adulte?

Paso 1: SIN información adicional

¿Qué hacemos?



Tamara dixit:

Va a medir 168 cm.

Tamara dixit:

Va a medir 168 cm. ... y, más o menos, fijate lo que tenés alrededor...

“Fijate lo que tenés alrededor” \equiv Estudiar datos

Estudiaremos datos - Learning

Datos: información observada en otros individuos.

Predicción: Con los datos hacemos una *cuenta* para poder predecir que ocurrirá con un nuevo sujeto.

Volvamos a nuestra pregunta

La pregunta: ¿Cuánto medirá al ser adult^e?

Paso 1: SIN información adicional

¿Qué hacemos?

Volvamos a nuestra pregunta

La pregunta: ¿Cuánto medirá al ser adult^e?

Paso 1: SIN información adicional

¿Qué hacemos?

¡Buscamos datos!

¿Cuánto medís?

Completemos la columna `altura` (en cm.) del documento compartido [► Planilla](#)

Predicción: ¿Qué cuenta hacemos?

Habemus datos. Y ahora...
¿Qué cuenta hacemos?



Predicción: ¿Qué cuenta hacemos?

Habemus datos. Y ahora...
¿Qué cuenta hacemos?



En este taller...

PROMEDIAMOS

Volvamos a nuestros datos y promediamos. [▶ Planilla](#)

Misma pregunta con información adicional

La pregunta: ¿Cuánto medirá al ser adulte?

Información Adicional: Será varón

Buscamos datos.

Volvamos a nuestra [Planilla](#), completemos la columna `genero`, con M, F y ...

¡PROMEDIEMOS!

Misma pregunta con información adicional

La pregunta: ¿Cuánto medirá al ser adulte?

Información Adicional: Será varón y la mamá es bajita.

Buscamos datos.

Volvamos a nuestra [Planilla](#), completemos la columna `contextura_madre`, con
bajita, mediana, alta y ...
¡PROMEDIEMOS!

Manos a la obra - Parte I: Calentando motores (15 minutos)

- Levantar los datos del archivo CSV que estan en `https://pilas.exp.dc.uba.ar/datos/alturas_n_50.csv`.
- Identificar el nombre de las columnas (variables) del archivo de datos.
- ¿Cuántos individuos conforman tu conjunto de observaciones?
- Calcular el promedio de las alturas de los hijos para predecir la altura de un nuevo individuo.
- ¿Con qué valor se puede predecir la altura de un nuevo individuo masculino?
- ¿Con qué valor se puede predecir la altura de un nuevo individuo masculino cuya madre es bajita?

¿Qué estuvimos haciendo?

- Resignificamos el promedio.
- De *medida de resumen* a **predictor**.
- Ideas simples y poderosas.
- La compu ayuda cuando hay muchos datos.

¿Qué hora es?

Misma pregunta con información adicional

La pregunta: ¿Cuánto medirá al ser adult^e?

Información Adicional: Será varón y la mamá mide 156.5 cm.

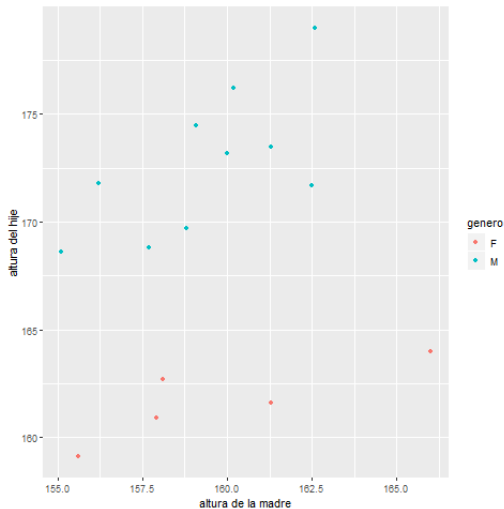
Datos: Incluimos la altura de la madre

Buscamos datos.

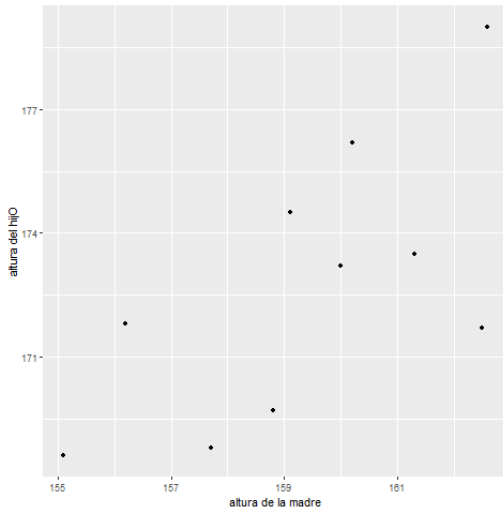
Completemos la columna `altura_madre` del documento compartido

► Planilla

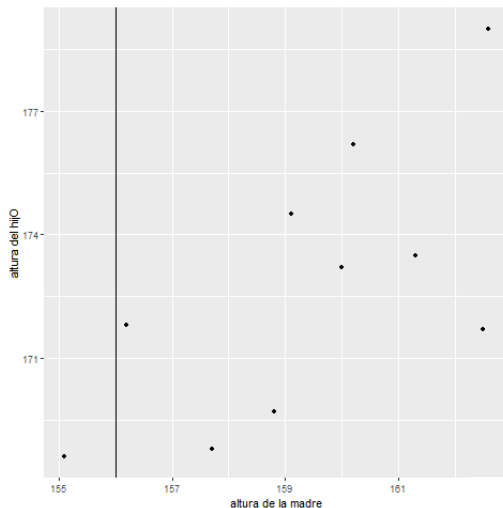
Posible Visualización (otros datos)- ¿Qué vemos?



Nos quedamos con los hijos: hijos varones



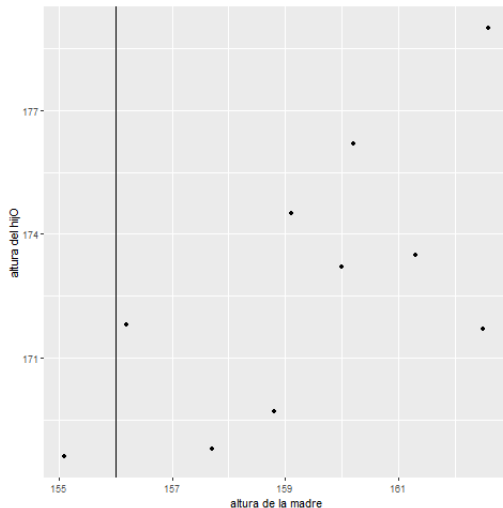
Nos quedamos con los varones y la madre mide 156.5 cm



¿Cómo podemos predecir la altura del hijo de esta mamá?

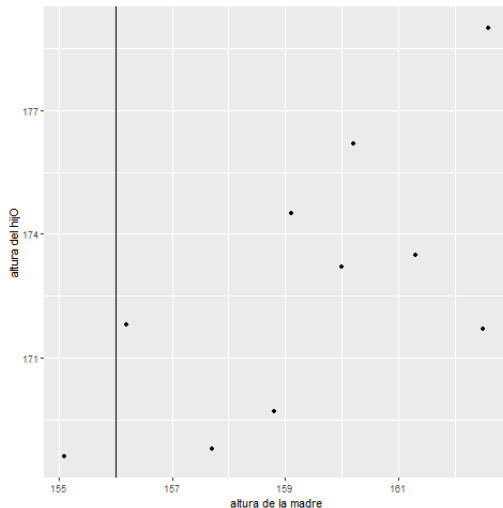


Nos quedamos con los varones y la madre mide 156.5 cm



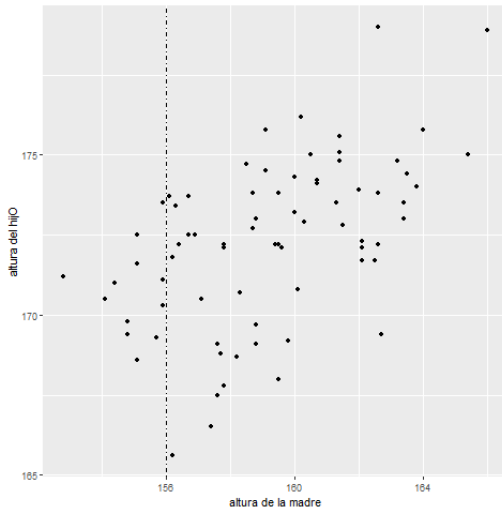
No toca! ¿Qué hacemos?

Nos quedamos con los varones y la madre mide 156.5 cm



No toca! ¿Qué hacemos?
Buscamos más datos

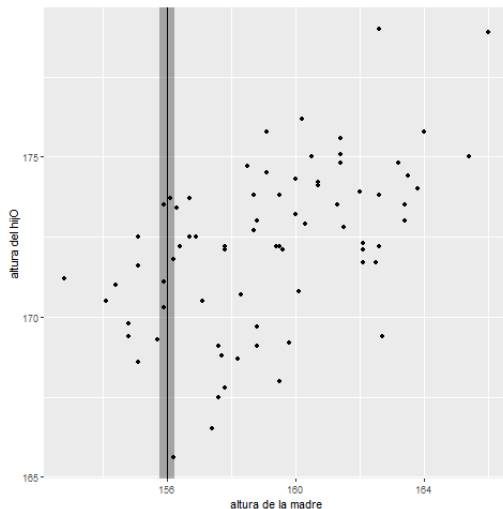
Más datos - sigue sin tocar



Silvana dixit:

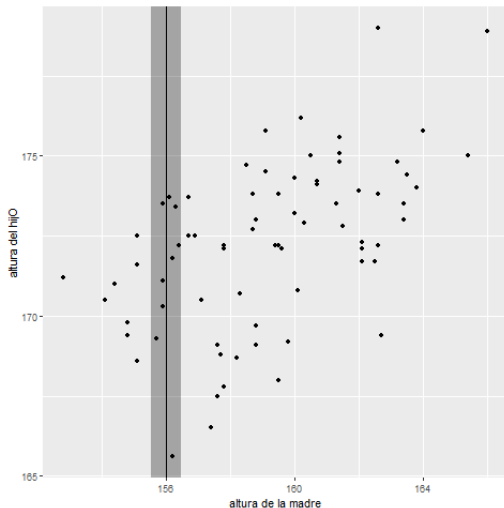
Hay unos puntitos que parece que pasan por la barrita, ¡pero no pasan! Para que pasen bien, habría que ensanchar la barrita.

Ensanchamos la barrita - Opción 1



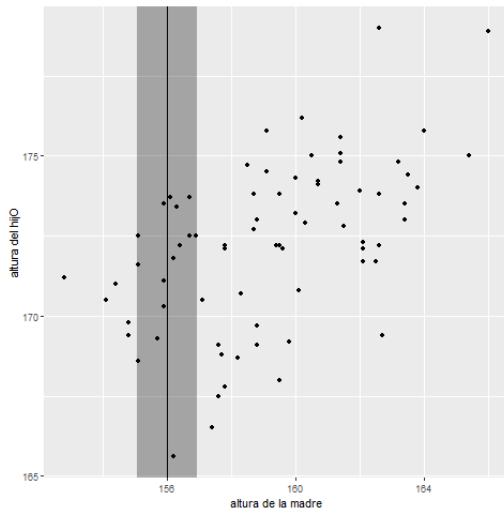
Ensanchamos la barrita y promediamos las alturas de los hijos correspondientes a los pares que caen adentro

Ensanchamos la barrita - Opción 2



Ensanchamos la barrita (más) y promediamos las alturas de los hijos de los pares que caen adentro

Ensanchamos la barrita - Opción 3



Ensanchamos la barrita (más aún) y promediamos las alturas de los hijos de los pares que caen adentro

¿Qué estuvimos haciendo? - Predecimos con promedios móviles

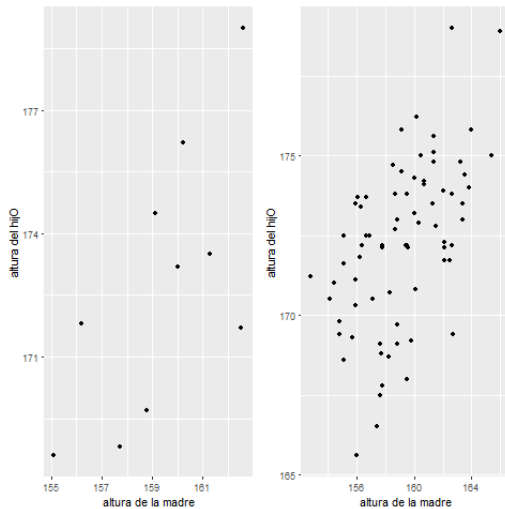
- 1 Determinamos la altura x_{nuevo} de la madre donde se quiere predecir (156.5 cm)
- 2 Elegimos un valor h de ventana para armar la barrita : h para derecha y h para izquierda.
- 3 Promediamos las alturas de los hijos correspondientes a los pares que caen adentro de la barrita con ventana de tamaño h ($\mp h$) centrada en x_{nuevo} .

Manos a la obra - Parte II: Mi primer promedio móvil (o local) (15 minutos)

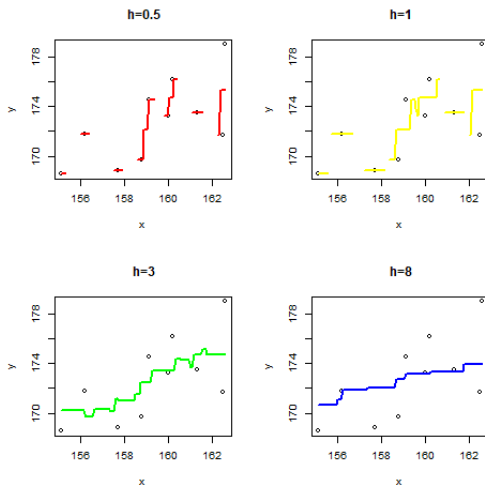
Nos quedamos con los datos con `genero == "M"`

- 1 Promedio móvil centrado en 156,5 con ventana de tamaño $h = 1$ (cm).
 - 1 Indique cuántos casos hay donde la madre registre una altura entre 155,5 y 157,5 cm, inclusive.
 - 2 Calcule el promedio de la altura de los hijos cuyas madres registran una altura entre 155,5 y 157,5 cm.

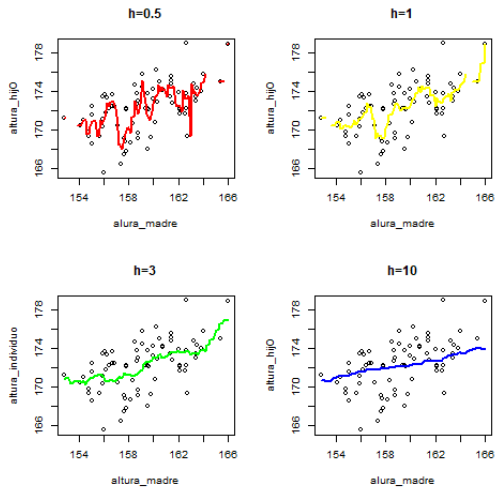
Pocos - Muchos



Promedios móviles con pocos datos



Promedios móviles con *muchos* datos



¿ Cómo elegimos el tamaño de la ventana?

Continuará

El jardín de senderos que se bifurcan

- 1 Elegir a las k - madres más cercanas y promediar la altura de sus hijos.
- 2 Calcular medianas en lugar de promedios (otro día).
- 3 ¿Y cuadrados mínimos?
- 4 ¡A trabajar!