

Nivelación - Estadística - TP - Complemento / Parte 2

Diego Dell'Era

Gracias por la devolución. Copio tu corrección:

95 percent confidence interval: 0.1767 0.6291

mean in group F mean in group M 11.15 10.75

(...) el intervalo de confianza no incluye al 0, eso quiere decir que sí hay diferencia de medias.
Estoy en lo correcto?

Sí, mala mía. Antes puse que no podía rechazar cómodamente la *null-hypothesis* porque el intervalo estaba casi todo alrededor del cero... pero es cierto, *no contiene el cero*, así que las medias de los dos grupos *son distintas*.

(Lo que sigue es lo que cambio en el TP original a partir del test de hipótesis.)

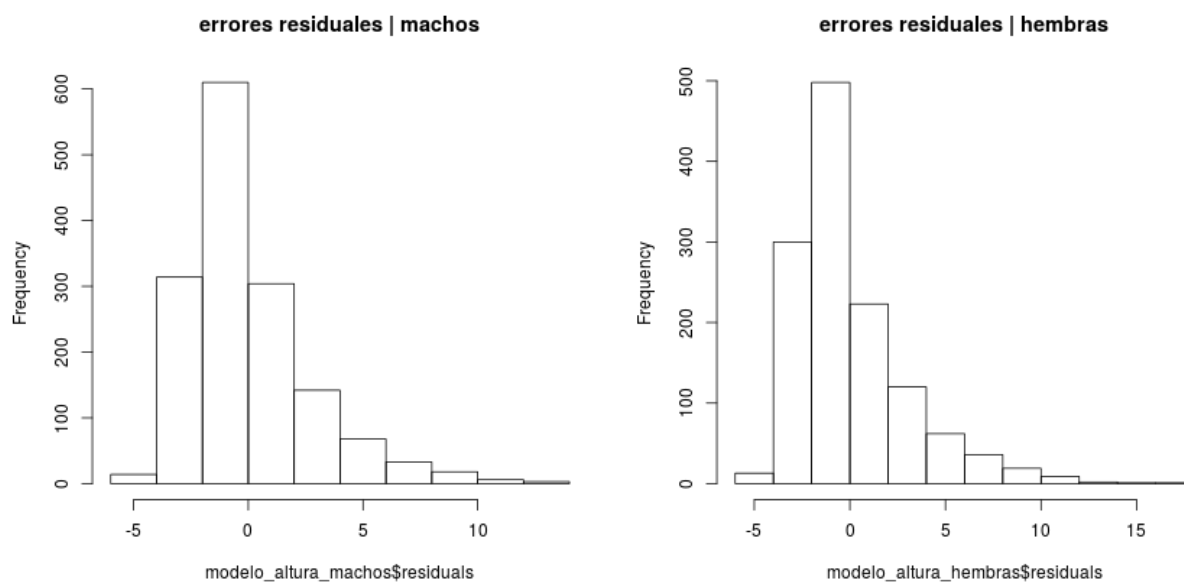
Conclusión modificada, entonces: los dos grupos comparados en el t-test (machos y hembras) se comportan con una pequeña diferencia con respecto a los Anillos que denotan su Edad.

Eso significa que hace falta otra modificación: no puedo aplicar un único modelo a todos los datos; tengo que armar un modelo por sexo.

```
modelo_altura_machos <- lm(Edad ~ Altura, data = machos)
modelo_altura_hembras <- lm(Edad ~ Altura, data = hembras)
```

La distribución de los errores residuales sigue siendo bastante parecida:

```
par(mfrow=c(1,2))
hist(modelo_altura_machos$residuals, main = "errores residuales | machos")
hist(modelo_altura_hembras$residuals, main = "errores residuales | hembras")
```



Si comparamos ahora el valor de *R-squared* (i.e., la variabilidad en la Edad que se puede explicar mediante la Altura), es un poquito mejor (i.e., un poco más cercano a 1) en el modelo de los machos que en el de las hembras:

```
summary(modelo_altura_machos)$r.squared
```

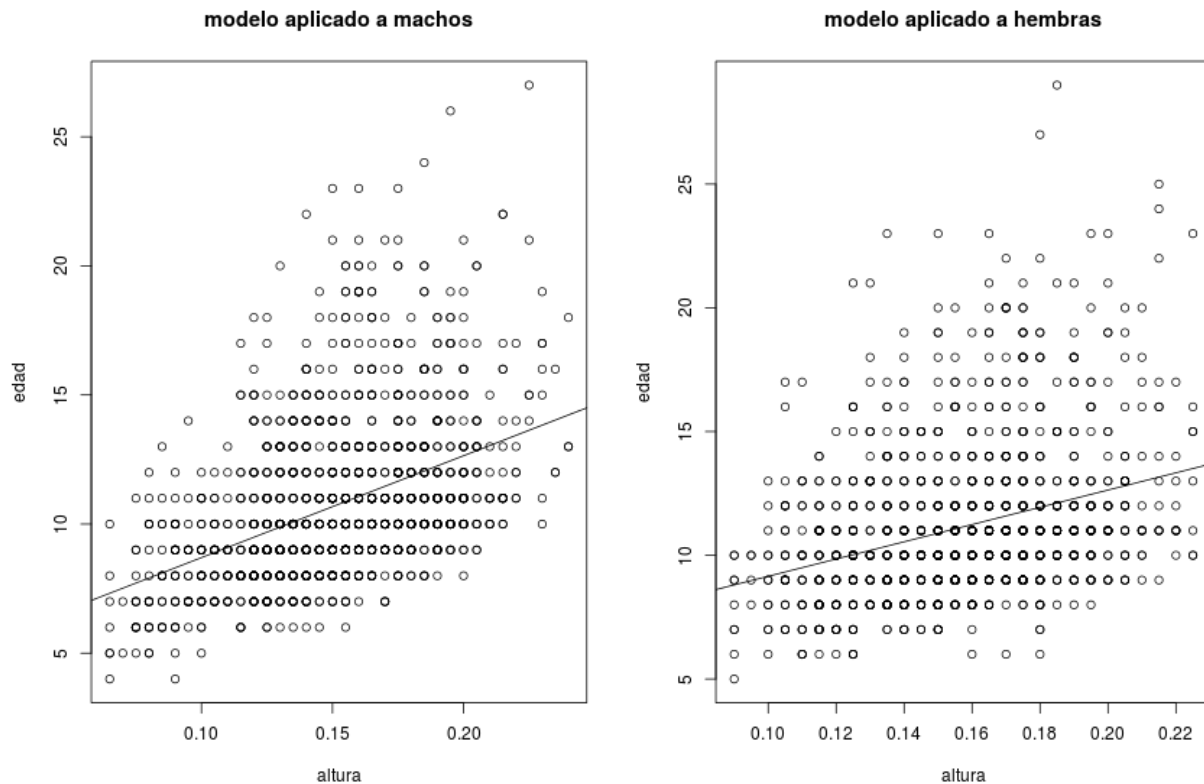
```
## [1] 0.1804
```

```
summary(modelo_altura_hembras)$r.squared
```

```
## [1] 0.09801
```

Si plotamos ambos modelos vemos que cambiaron un poco las pendientes de las respectivas rectas...

```
par(mfrow=c(1,2))
plot(machos$Altura, machos$Edad, xlab="altura", ylab="edad", main="modelo aplicado a machos")
abline(modelo_altura_machos)
plot(hembras$Altura, hembras$Edad, xlab="altura", ylab="edad", main="modelo aplicado a hembras")
abline(modelo_altura_hembras)
```



... pero en ambos casos sigue pasando lo mismo que pasaba con el modelo original: quedan muchos dispersos por encima. La conclusión se mantiene: predecir Edad usando Altura no es una buena aproximación :(