Estadística y Ajustes

Lucas Finazzi

Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Septiembre 2021

universidad de buenos aires - exactas
departamento de física

Qué error se pone a un promedio?

Dada una serie de puntos con una desviación estándar σ y valor esperado μ,

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i \quad ,$$

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$$

son buenos estimadores de ambas cantidades.

Qué error se pone a un promedio?

Sin embargo, el error en el promedio es más chico que la desviación estándar! Que tanto más chico?

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{N}} \quad .$$

Y si los puntos tienen errores diferentes?

Si cada punto tiene un error distinto, esto se puede generalizar y se puede escribir

$$\bar{Y} = \frac{\sum_{i=1}^{N} \frac{Y_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \quad ,$$

$$\sigma_{\bar{Y}}^2 = \frac{1}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \quad .$$

Por ejemplo, esto puede usarse para combinar mediciones de distintos experimentos de la misma cantidad.

Algunas cosas sobre ajustes

- Todos los puntos no van a estar sobre la curva ajustada.
- No todas las barras de error deben contener al ajuste! Si se tienen errores gaussianos, esperamos que solamente el 68 % de los puntos con sus barras de error contengan el ajuste.
- Un punto a más de 3σ del ajuste es muy raro. Si los errores son gaussianos, ocurre 1 cada 370 veces!
- **Es importante siempre ver el gráfico!**

Qué no es el R^2

El R^2 **no** es un parámetro de bondad de ajuste! Sólo me indica qué tan sensibles son los cambios entre dos variables. En el caso lineal,

$$R^2 = \frac{\Big(\sum_i (X_i - \bar{X})(Y_i - \bar{Y})\Big)^2}{\sum_i (X_i - \bar{X})^2 \sum_i (Y_i - \bar{Y})^2} \quad .$$

No depende del ajuste! Solo de los datos. Sin embargo, puede ser útil como herramienta en ciertos casos.

Parámetros de bondad de ajuste: El estadístico χ^2

El estadístico χ^2_ν me habla de qué tan lejos están los puntos del ajuste realizado. Se define como

$$\chi^2_\nu = \sum_{i=1}^{N} \left(\frac{Y_i - f(X_i)}{\sigma_i} \right)^2 \quad ,$$

donde ν es la cantidad de grados de libertad. El valor esperado es $\nu \pm \sqrt{2\nu}$. La cantidad de grados de libertad es la cantidad de puntos ajustados menos la cantidad de parámetros a ajustar, i.e. $\nu = N - p$.

Test de hipótesis: El p-valor

En el contexto actual, el p-valor se define como

La probabilidad de haber medido un χ^2 *como el que medí o más extremo tal que mi hipótesis inicial es verdadera.*

El p-valor **NO** es

La probabilidad de que mi teoría sea correcta.

¡Muchas Gracias!