

A photograph of two lion cubs in a savanna setting. One cub is on the left, facing right, and the other is on the right, facing left. They are both on their hind legs, reaching towards each other with their front paws. The background is a blurred green field with some tall grass. The text 'AST0212 – 2016-1' is overlaid in the center in a large, bold, yellow font.

AST0212 – 2016-1

Introducción al análisis de datos

Instituto de Astrofísica

Facultad de Física

Pontificia Universidad Católica de Chile

Nuestro Semestre 2016-1

| AST0212 | | | | C0 ✓ | | |
|------------------------|--------|--------------|-----------|--------------|--------------|--------------------------------|
| Sunday | Monday | Tuesday | Wednesday | Thursday | Friday | Saturday |
| 6 Mar 2016 Semana 1 | | | | | C1 ✓ | |
| 13 Semana 2 | TL1 | TM1 | | | C2 ✓ | ← Control 1 Reparto Tarea 1 |
| 20 Semana 3 | TL2 | TM2 | | | Feriado | |
| 27 Semana 4 | TL3 | TM3 | | | C3 ✓ | |
| 3 Semana 5 | TL4 | TM4 | | | C4 ✓ | |
| 10 Semana 6 | TL5 | TM5 | | | C5 ✓ | ← Control 2 |
| 17 Semana 7 | TL6 | TM6 | | | C6 ✓ – SM1 ✗ | ← Reparto T2 |
| 24 Semana 8 | TL7 | ← Entrega T1 | | | C7 ✓ – SM2 ✗ | |
| 1 May Semana 9 | TL8 | TM8 | | | C8 ✓ – SM3 ✗ | |
| 8 Semana 10 | TL9 | Entrega T2 → | | Reparto T3 → | C9 – SM4 | ← Control 3 |
| 15 Semana 11 | TL10 | TM10 | | | C10 | |
| 22 Semana 12 | TL11 | Entrega T3 → | | | C11 | |
| 29 Semana 13 | TL12 | TM12 | 1 Jun | 2 | Feriado | |
| 5 Semana 14 | TL13 | TM13 | 8 | 9 | C12 | |
| 12 Semana 15 | TL14 | TM14 | 15 | 16 | C13 | |
| 19 | | | 22 | 23 | | |
| 26 | | | 29 | 30 | | |
| 3 Semana 16 | | | 6 | 7 | | |
| 10 Semana 17 | | | 13 | 14 | | |
| 17 Semana 18 | | | 20 | 21 | | |
| 24 Semana 19 | | | 27 | 28 | | |
| 31 Semana 20 | | | 4 | 5 | | |
| 7 Semana 21 | | | 11 | 12 | | |
| 14 Semana 22 | | | 18 | 19 | | |
| 21 Semana 23 | | | 25 | 26 | | |
| 28 Semana 24 | | | 1 Jun | 2 | | |
| 5 Semana 25 | | | 8 | 9 | | |
| 12 Semana 26 | | | 15 | 16 | | |
| 19 Semana 27 | | | 22 | 23 | | |
| 26 Semana 28 | | | 29 | 30 | | |
| 3 Semana 29 | | | 6 | 7 | | |
| 10 Semana 30 | | | 13 | 14 | | |
| 17 Semana 31 | | | 20 | 21 | | |
| 24 Semana 32 | | | 27 | 28 | | |
| 31 Semana 33 | | | 4 | 5 | | |
| 7 Semana 34 | | | 11 | 12 | | |
| 14 Semana 35 | | | 18 | 19 | | |
| 21 Semana 36 | | | 25 | 26 | | |
| 28 Semana 37 | | | 1 Jun | 2 | | |
| 5 Semana 38 | | | 8 | 9 | | |
| 12 Semana 39 | | | 15 | 16 | | |
| 19 Semana 40 | | | 22 | 23 | | |
| 26 Semana 41 | | | 29 | 30 | | |
| 3 Semana 42 | | | 6 | 7 | | |
| 10 Semana 43 | | | 13 | 14 | | |
| 17 Semana 44 | | | 20 | 21 | | |
| 24 Semana 45 | | | 27 | 28 | | |
| 31 Semana 46 | | | 4 | 5 | | |
| 7 Semana 47 | | | 11 | 12 | | |
| 14 Semana 48 | | | 18 | 19 | | |
| 21 Semana 49 | | | 25 | 26 | | |
| 28 Semana 50 | | | 1 Jun | 2 | | |
| 5 Semana 51 | | | 8 | 9 | | |
| 12 Semana 52 | | | 15 | 16 | | |
| 19 Semana 53 | | | 22 | 23 | | |
| 26 Semana 54 | | | 29 | 30 | | |
| 3 Semana 55 | | | 6 | 7 | | |
| 10 Semana 56 | | | 13 | 14 | | |
| 17 Semana 57 | | | 20 | 21 | | |
| 24 Semana 58 | | | 27 | 28 | | |
| 31 Semana 59 | | | 4 | 5 | | |
| 7 Semana 60 | | | 11 | 12 | | |
| 14 Semana 61 | | | 18 | 19 | | |
| 21 Semana 62 | | | 25 | 26 | | |
| 28 Semana 63 | | | 1 Jun | 2 | | |
| 5 Semana 64 | | | 8 | 9 | | |
| 12 Semana 65 | | | 15 | 16 | | |
| 19 Semana 66 | | | 22 | 23 | | |
| 26 Semana 67 | | | 29 | 30 | | |
| 3 Semana 68 | | | 6 | 7 | | |
| 10 Semana 69 | | | 13 | 14 | | |
| 17 Semana 70 | | | 20 | 21 | | |
| 24 Semana 71 | | | 27 | 28 | | |
| 31 Semana 72 | | | 4 | 5 | | |
| 7 Semana 73 | | | 11 | 12 | | |
| 14 Semana 74 | | | 18 | 19 | | |
| 21 Semana 75 | | | 25 | 26 | | |
| 28 Semana 76 | | | 1 Jun | 2 | | |
| 5 Semana 77 | | | 8 | 9 | | |
| 12 Semana 78 | | | 15 | 16 | | |
| 19 Semana 79 | | | 22 | 23 | | |
| 26 Semana 80 | | | 29 | 30 | | |
| 3 Semana 81 | | | 6 | 7 | | |
| 10 Semana 82 | | | 13 | 14 | | |
| 17 Semana 83 | | | 20 | 21 | | |
| 24 Semana 84 | | | 27 | 28 | | |
| 31 Semana 85 | | | 4 | 5 | | |
| 7 Semana 86 | | | 11 | 12 | | |
| 14 Semana 87 | | | 18 | 19 | | |
| 21 Semana 88 | | | 25 | 26 | | |
| 28 Semana 89 | | | 1 Jun | 2 | | |
| 5 Semana 90 | | | 8 | 9 | | |
| 12 Semana 91 | | | 15 | 16 | | |
| 19 Semana 92 | | | 22 | 23 | | |
| 26 Semana 93 | | | 29 | 30 | | |
| 3 Semana 94 | | | 6 | 7 | | |
| 10 Semana 95 | | | 13 | 14 | | |
| 17 Semana 96 | | | 20 | 21 | | |
| 24 Semana 97 | | | 27 | 28 | | |
| 31 Semana 98 | | | 4 | 5 | | |
| 7 Semana 99 | | | 11 | 12 | | |
| 14 Semana 100 | | | 18 | 19 | | |
| 21 Semana 101 | | | 25 | 26 | | |
| 28 Semana 102 | | | 1 Jun | 2 | | |
| 5 Semana 103 | | | 8 | 9 | | |
| 12 Semana 104 | | | 15 | 16 | | |
| 19 Semana 105 | | | 22 | 23 | | |
| 26 Semana 106 | | | 29 | 30 | | |
| 3 Semana 107 | | | 6 | 7 | | |
| 10 Semana 108 | | | 13 | 14 | | |
| 17 Semana 109 | | | 20 | 21 | | |
| 24 Semana 110 | | | 27 | 28 | | |
| 31 Semana 111 | | | 4 | 5 | | |
| 7 Semana 112 | | | 11 | 12 | | |
| 14 Semana 113 | | | 18 | 19 | | |
| 21 Semana 114 | | | 25 | 26 | | |
| 28 Semana 115 | | | 1 Jun | 2 | | |
| 5 Semana 116 | | | 8 | 9 | | |
| 12 Semana 117 | | | 15 | 16 | | |
| 19 Semana 118 | | | 22 | 23 | | |
| 26 Semana 119 | | | 29 | 30 | | |
| 3 Semana 120 | | | 6 | 7 | | |
| 10 Semana 121 | | | 13 | 14 | | |
| 17 Semana 122 | | | 20 | 21 | | |
| 24 Semana 123 | | | 27 | 28 | | |
| 31 Semana 124 | | | 4 | 5 | | |
| 7 Semana 125 | | | 11 | 12 | | |
| 14 Semana 126 | | | 18 | 19 | | |
| 21 Semana 127 | | | 25 | 26 | | |
| 28 Semana 128 | | | 1 Jun | 2 | | |
| 5 Semana 129 | | | 8 | 9 | | |
| 12 Semana 130 | | | 15 | 16 | | |
| 19 Semana 131 | | | 22 | 23 | | |
| 26 Semana 132 | | | 29 | 30 | | |
| 3 Semana 133 | | | 6 | 7 | | |
| 10 Semana 134 | | | 13 | 14 | | |
| 17 Semana 135 | | | 20 | 21 | | |
| 24 Semana 136 | | | 27 | 28 | | |
| 31 Semana 137 | | | 4 | 5 | | |
| 7 Semana 138 | | | 11 | 12 | | |
| 14 Semana 139 | | | 18 | 19 | | |
| 21 Semana 140 | | | 25 | 26 | | |
| 28 Semana 141 | | | 1 Jun | 2 | | |
| 5 Semana 142 | | | 8 | 9 | | |
| 12 Semana 143 | | | 15 | 16 | | |
| 19 Semana 144 | | | 22 | 23 | | |
| 26 Semana 145 | | | 29 | 30 | | |
| 3 Semana 146 | | | 6 | 7 | | |
| 10 Semana 147 | | | 13 | 14 | | |
| 17 Semana 148 | | | 20 | 21 | | |
| 24 Semana 149 | | | 27 | 28 | | |
| 31 Semana 150 | | | 4 | 5 | | |
| 7 Semana 151 | | | 11 | 12 | | |
| 14 Semana 152 | | | 18 | 19 | | |
| 21 Semana 153 | | | 25 | 26 | | |
| 28 Semana 154 | | | 1 Jun | 2 | | |
| 5 Semana 155 | | | 8 | 9 | | |
| 12 Semana 156 | | | 15 | 16 | | |
| 19 Semana 157 | | | 22 | 23 | | |
| 26 Semana 158 | | | 29 | 30 | | |
| 3 Semana 159 | | | 6 | 7 | | |
| 10 Semana 160 | | | 13 | 14 | | |
| 17 Semana 161 | | | 20 | 21 | | |
| 24 Semana 162 | | | 27 | 28 | | |
| 31 Semana 163 | | | 4 | 5 | | |
| 7 Semana 164 | | | 11 | 12 | | |
| 14 Semana 165 | | | 18 | 19 | | |
| 21 Semana 166 | | | 25 | 26 | | |
| 28 Semana 167 | | | 1 Jun | 2 | | |
| 5 Semana 168 | | | 8 | 9 | | |
| 12 Semana 169 | | | 15 | 16 | | |
| 19 Semana 170 | | | 22 | 23 | | |
| 26 Semana 171 | | | 29 | 30 | | |
| 3 Semana 172 | | | 6 | 7 | | |
| 10 Semana 173 | | | 13 | 14 | | |
| 17 Semana 174 | | | 20 | 21 | | |
| 24 Semana 175 | | | 27 | 28 | | |
| 31 Semana 176 | | | 4 | 5 | | |
| 7 Semana 177 | | | 11 | 12 | | |
| 14 Semana 178 | | | 18 | 19 | | |
| 21 Semana 179 | | | 25 | 26 | | |
| 28 Semana 180 | | | 1 Jun | 2 | | |
| 5 Semana 181 | | | 8 | 9 | | |
| 12 Semana 182 | | | 15 | 16 | | |
| 19 Semana 183 | | | 22 | 23 | | |
| 26 Semana 184 | | | 29 | 30 | | |
| 3 Semana 185 | | | 6 | 7 | | |
| 10 Semana 186 | | | 13 | 14 | | |
| 17 Semana 187 | | | 20 | 21 | | |
| 24 Semana 188 | | | 27 | 28 | | |
| 31 Semana 189 | | | 4 | 5 | | |
| 7 Semana 190 | | | 11 | 12 | | |
| 14 Semana 191 | | | 18 | 19 | | |
| 21 Semana 192 | | | 25 | 26 | | |
| 28 Semana 193 | | | 1 Jun | 2 | | |
| 5 Semana 194 | | | 8 | 9 | | |
| 12 Semana 195 | | | 15 | 16 | | |
| 19 Semana 196 | | | 22 | 23 | | |
| 26 Semana 197 | | | 29 | 30 | | |
| 3 Semana 198 | | | 6 | 7 | | |
| 10 Semana 199 | | | 13 | 14 | | |
| 17 Semana 200 | | | 20 | 21 | | |
| 24 Semana 201 | | | 27 | 28 | | |
| 31 Semana 202 | | | 4 | 5 | | |
| 7 Semana 203 | | | 11 | 12 | | |
| 14 Semana 204 | | | 18 | 19 | | |
| 21 Semana 205 | | | 25 | 26 | | |
| 28 Semana 206 | | | 1 Jun | 2 | | |
| 5 Semana 207 | | | 8 | 9 | | |
| 12 Semana 208 | | | 15 | 16 | | |
| 19 Semana 209 | | | 22 | 23 | | |
| 26 Semana 210 | | | 29 | 30 | | |
| 3 Semana 211 | | | 6 | 7 | | |
| 10 Semana 212 | | | 13 | 14 | | |
| 17 Semana 213 | | | 20 | 21 | | |
| 24 Semana 214 | | | 27 | 28 | | |
| 31 Semana 215 | | | 4 | 5 | | |
| 7 Semana 216 | | | 11 | 12 | | |
| 14 Semana 217 | | | 18 | 19 | | |
| 21 Semana 218 | | | 25 | 26 | | |
| 28 Semana 219 | | | 1 Jun | 2 | | |
| 5 Semana 220 | | | 8 | 9 | | |
| 12 Semana 221 | | | 15 | 16 | | |
| 19 Semana 222 | | | 22 | 23 | | |
| 26 Semana 223 | | | 29 | 30 | | |
| 3 Semana 224 | | | 6 | 7 | | |
| 10 Semana 225 | | | 13 | 14 | | |
| 17 Semana 226 | | | 20 | 21 | | |
| 24 Semana 227 | | | 27 | 28 | | |
| 31 Semana 228 | | | 4 | 5 | | |
| 7 Semana 229 | | | 11 | 12 | | |
| 14 Semana 230 | | | 18 | 19 | | |
| 21 Semana 231 | | | 25 | 26 | | |
| 28 Semana 232 | | | 1 Jun | 2 | | |
| 5 Semana 233 | | | 8 | 9 | | |
| 12 Semana 234 | | | 15 | 16 | | |
| 19 Semana 235 | | | 22 | 23 | | |
| 26 Semana 236 | | | 29 | 30 | | |
| 3 Semana 237 | | | 6 | 7 | | |
| 10 Semana 238 | | | 13 | 14 | | |
| 17 Semana 239 | | | 20 | 21 | | |
| 24 Semana 240 | | | 27 | 28 | | |
| 31 Semana 241 | | | 4 | 5 | | |
| 7 Semana 242 | | | 11 | 12 | | |
| 14 Semana 243 | | | 18 | 19 | | |
| 21 Semana 244 | | | 25 | 26 | | |
| 28 Semana 245 | | | 1 Jun | 2 | | |
| 5 Semana 246 | | | 8 | 9 | | |
| 12 Semana 247 | | | 15 | 16 | | |
| 19 Semana 248 | | | 22 | 23 | | |
| 26 Semana 249 | | | 29 | 30 | | |
| 3 Semana 250 | | | 6 | 7 | | |
| 10 Semana 251 | | | 13 | 14 | | |
| 17 Semana 252 | | | 20 | 21 | | |
| 24 Semana 253 | | | 27 | 28 | | |

- 1) ¿Qué es un histograma? ¿Cómo se construye? ¿Cómo puede caracterizarse?
- 2) ¿Qué es una FDP?
- 3) ¿Cuál es la relación entre la FDP de una cierta variable y el histograma de valores que medimos para esta misma variable?
- 4) Dado un conjunto de medidas (datos) de valores directamente comparables entre sí, defina el valor medio, la mediana, la moda, y la dispersión.
- 5) ¿En qué clase de experimentos la dispersión de una variable observada proporciona una medida de la incerteza en la medición?
- 6) ¿Cuál es la diferencia entre una incerteza en la precisión y una en la exactitud?
- 7) Si queremos conocer una variable t , que no podemos medir, pero que se relaciona con otras variables x, y, z , que sí podemos medir directamente, por la ecuación

$$t = f(x, y, z)$$

- a. Dadas K medidas de x , N medidas de y , y M medidas de z , explique cómo haría para calcular \bar{t} y σ_t .
- b. Puede imaginar una estrategia diferente para calcular \bar{t} y σ_t para el caso de disponer de N medidas de x, y, z ? (i.e. la misma cantidad de medidas en cada variable)

Clase previa (Clase 7):

REPASO

1. Repaso de temas críticos de la clase previa ✓
 1. Correlación.
 2. Incerteza de parámetros en la correlación lineal.
 3. Corrección de error sistemático. Extrapolación.
2. Significación de diferencia en media y ~~varianza~~ ✓
3. Coeficiente de correlación. ✓

Esta clase (Clase 8):

1. Repaso de temas críticos de la clase previa
 1. Correlación.
 2. Incerteza de parámetros en la correlación lineal.
 3. Corrección de error sistemático. Extrapolación.
2. Coeficiente de correlación.
3. Significación de diferencia en media y varianza

Comparación de dos distribuciones observadas

REPASO

$$\chi^2 = \sum_{j=1}^M \frac{(n_{1,j} - n_{2,j})^2}{n_{1,j} + n_{2,j}}, \text{ donde } M \text{ es el número de bins.}$$

La FDP de este χ^2 es la misma que mostré antes. ¿Qué es ν ahora? Si los datos son recogidos de forma tal que la suma de n_1 es necesariamente igual a la de n_2 tendremos que el número de grados de libertad es $\nu = M - 1$ (el caso usual). Si este requerimiento no existe, entonces $\nu = M$.

Ejemplo: Un observador de aves que desea comparar dos años de observaciones, tomando un bin por cada especie.

1: Base de datos es los 1000 primeros pájaros que observa cada año ($\nu = M - 1$)

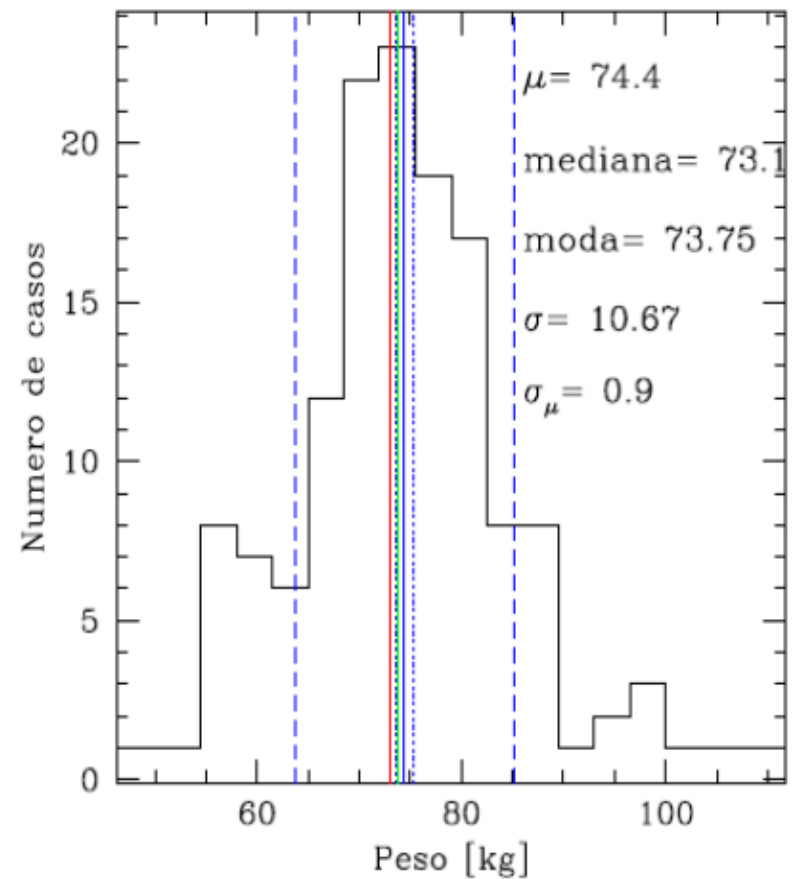
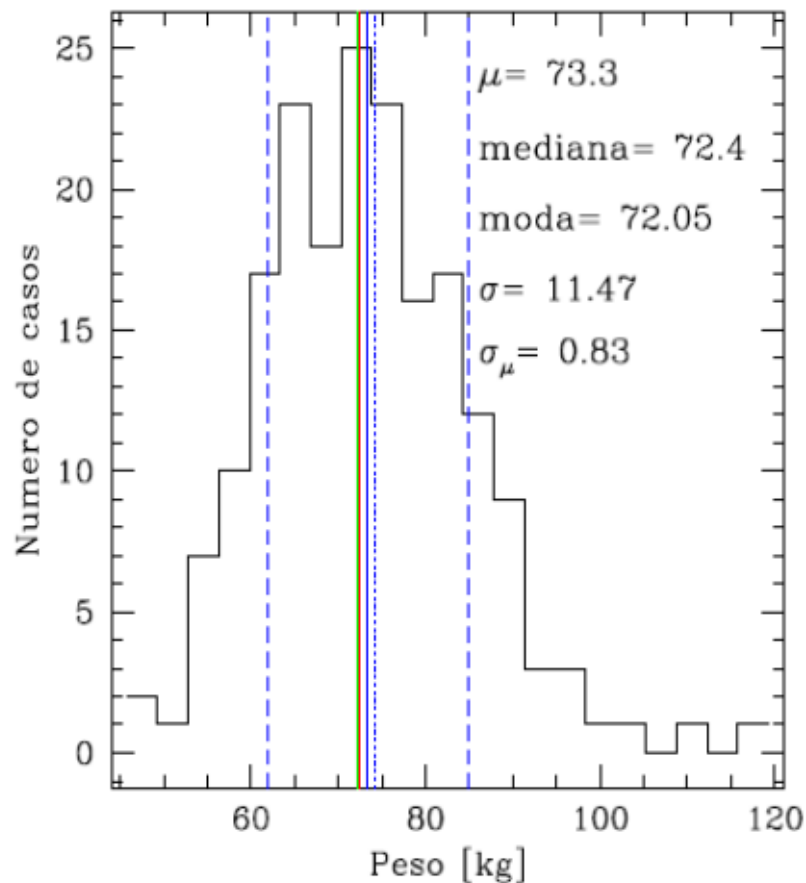
2: Base de datos es todos los pájaros que vio en un número de días al azar, siendo el número de días el mismo en los dos años ($\nu = M$).

En el segundo caso puede comparar los totales. *Ese es el grado de libertad adicional.*

Comparación de dos distribuciones observadas: Significación de la diferencia de promedios

¿REPASO?

G12345678-E-M-AC-peso.dat ; $N_T = 190$; Bin = 3.5 kg G12345678-E-M-DC-peso.dat ; $N_T = 141$; Bin = 3.5 kg



Comparación de dos distribuciones observadas:

Significación de la diferencia de promedios

REPASO

Dados:

$$\bar{x}_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} x_{1,i}$$

$$\sigma_1 = \frac{1}{N_1 - 1} \sum_{i=1}^{N_1} x_{1,i}$$

$$\bar{x}_2 = \frac{1}{N_2} \sum_{i=1}^{N_2} x_{2,i}$$

$$\sigma_2 = \frac{1}{N_2 - 1} \sum_{i=1}^{N_2} x_{2,i}$$

Tendremos el error del promedio:

¡Permite la estrategia de aumentar la muestra!

$$\sigma_{\bar{x}_1} = \frac{\sigma_1}{\sqrt{N_1}}$$

(Se obtienen de aplicar propagación de errores a las definiciones de \bar{x}_1 y \bar{x}_2 .)

$$\sigma_{\bar{x}_2} = \frac{\sigma_2}{\sqrt{N_2}}$$

Con estos elementos podemos construir el estimador t , con $\nu = N_1 + N_2 - 2$ grados de libertad:

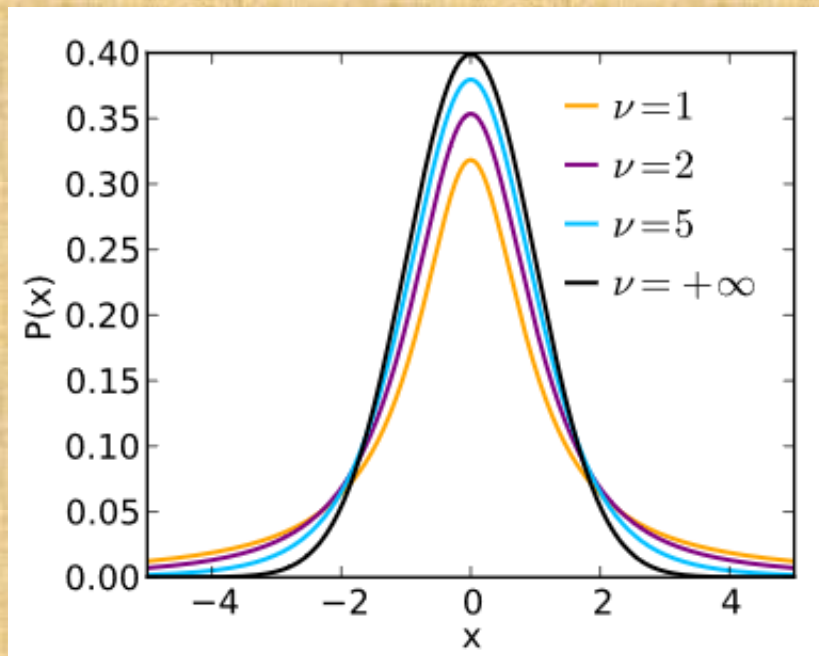
$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_D}$$

donde

$$S_D = \sqrt{\frac{\sum_{N_1} (x_{1,i} - \bar{x}_1)^2 + \sum_{N_2} (x_{2,i} - \bar{x}_2)^2}{N_1 + N_2 - 2} \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}$$

S_D es el error estándar de la diferencia de promedios. t tiene FDP tipo t – *Student*.

Distribución *t* de Student



La FDP de t , $A(t|\nu)$, denota la probabilidad de que t sea, por azar, menor que el valor medido si los promedios \bar{x}_1 y \bar{x}_2 son realmente iguales. Un valor grande (por ejemplo 0.99) indica una *alta chance de medir un valor menor que el observado si $\bar{x}_1 = \bar{x}_2$* . Esto es una indicación de que los promedios muy probablemente no sean los mismos. El valor complementario $1 - A(t|\nu)$ es la probabilidad de medir un valor tan grande como t si $\bar{x}_1 = \bar{x}_2$ (0.01 en el caso previo).

$$A(t|\nu) = \int_{-t}^t \frac{1}{\nu^{\frac{1}{2}} B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} dx = \int_{-t}^t P(x) dx = 1 - I_{\frac{\nu}{\nu+t^2}}\left(\frac{\nu}{2}, \frac{1}{2}\right)$$

Donde $B(a, b)$ es la función Beta, e $I_x(a, b)$ es la función Beta incompleta (en este caso para $x = \frac{\nu}{\nu+t^2}$, $a = \nu/2$ y $b = 1/2$).

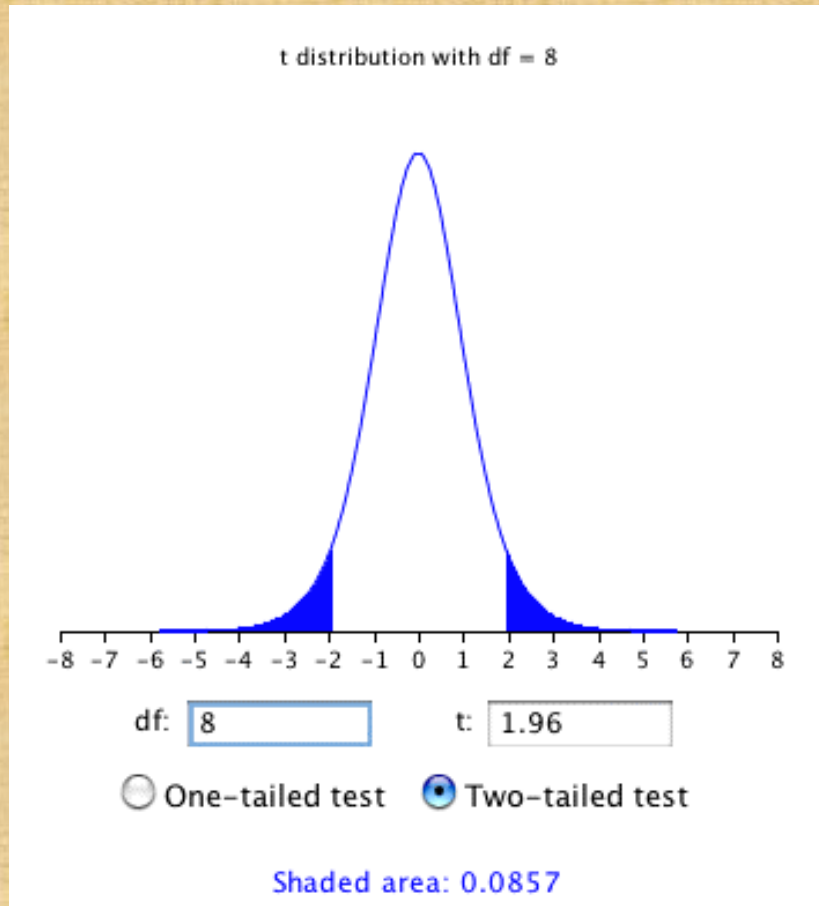
Hay calculadores on-line para estas funciones, por ejemplo para $1 - A(t|\nu)$:

http://onlinestatbook.com/2/calculators/t_dist.html

Distribución *t* de Student

$$A(t|\nu) = \frac{1}{\nu^{\frac{1}{2}} B(\frac{1}{2}, \frac{\nu}{2})} \int_{-t}^t \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} dx$$

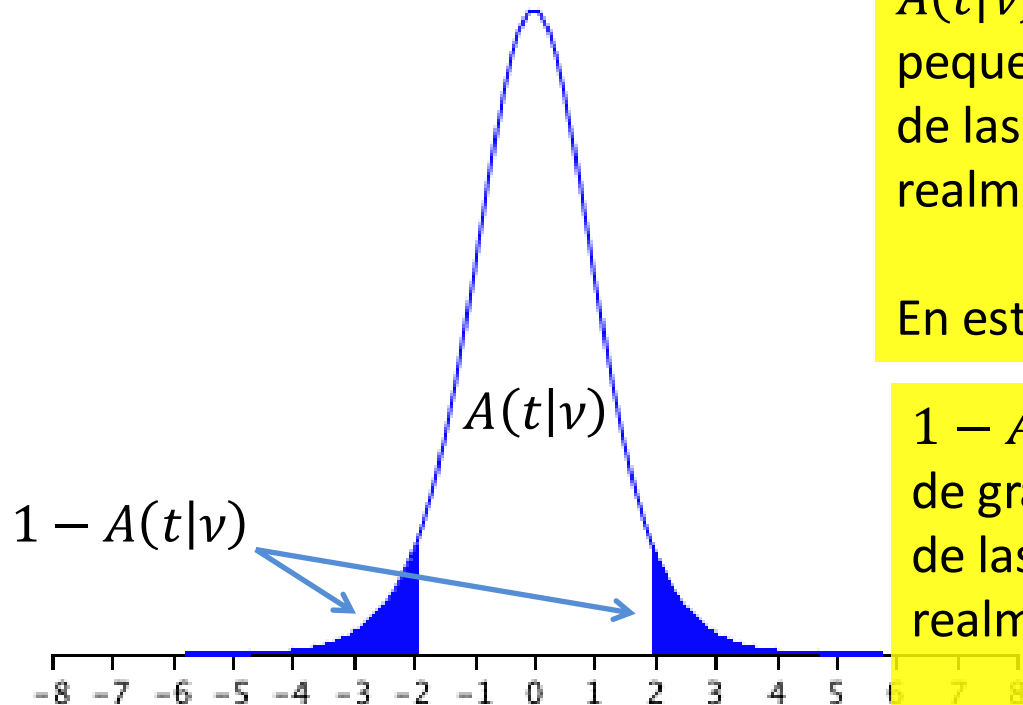
La FDP *t* de Student es en realidad una FDP acumulativa (la integral entre $-t$ y t). Lo que estaba graficado en la lámina previa era el integrando.



Este gráfico muestra $A(t|\nu)$ como área blanca bajo la línea azul, y $1 - A(t|\nu)$ como área azul en los extremos derecho e izquierdo de la distribución. Por simetría, debemos considerar ambas colas (ya que el orden en que hacemos la resta en la definición de t es arbitrario. La función $1 - A(t|\nu)$ incorpora las dos colas naturalmente.

Interpretación del *t de Student*

t distribution with df = 8



df:

8

t:

1.96



One-tailed test



Two-tailed test

Shaded area: 0.0857

$A(t|v)$: Probabilidad de que t sea así de pequeño por azar, cuando los promedios de las dos distribuciones comparadas son realmente iguales.

En este ejemplo: $A(1.96|8) = 0.9143$

$1 - A(t|v)$: Probabilidad de que t sea así de grande por azar, cuando los promedios de las dos distribuciones comparadas son realmente iguales.

En el ejemplo: $1 - A(1.96|8) = 0.0857$

Coeficiente de correlación

REPASO

Partamos con un recordatorio de las ecuaciones de ajuste lineal de cuadrados mínimos:

$$y_i = ax_i + b$$

$$\sum_{i=1}^N \frac{y_i}{\sigma_i^2} = b \sum_{i=1}^N \frac{1}{\sigma_i^2} + a \sum_{i=1}^N \frac{x_i}{\sigma_i^2}$$

$$\sum_{i=1}^N \frac{x_i y_i}{\sigma_i^2} = b \sum_{i=1}^N \frac{x_i}{\sigma_i^2} + a \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2}$$

$$\chi^2 = \sum_{i=1}^N \left(\frac{y_i - (ax_i + b)}{\sigma_i} \right)^2$$
$$\Delta = \sum_{i=1}^N \frac{1}{\sigma_i^2} \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^N \frac{x_i}{\sigma_i^2} \right)^2$$

$$a = \frac{1}{\Delta} \left(\sum_{i=1}^N \frac{1}{\sigma_i^2} \sum_{i=1}^N \frac{x_i y_i}{\sigma_i^2} - \sum_{i=1}^N \frac{x_i}{\sigma_i^2} \sum_{i=1}^N \frac{y_i}{\sigma_i^2} \right)$$

$$b = \frac{1}{\Delta} \left(\sum_{i=1}^N \frac{x_i^2}{\sigma_i^2} \sum_{i=1}^N \frac{y_i}{\sigma_i^2} - \sum_{i=1}^N \frac{x_i}{\sigma_i^2} \sum_{i=1}^N \frac{x_i y_i}{\sigma_i^2} \right)$$

Simplifiquemos para un caso sin σ (es idéntico a imaginar $\sigma = 1$)

Coeficiente de correlación (caso sin REPASO)

¿Tiene sentido la correlación $y_i = ax_i + b$? Prestemos atención a la pendiente.

$$a = \frac{1}{\Delta} \left(N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i \right) = \frac{\Delta_s}{\Delta}$$
$$\Delta = N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2$$
$$x_i = a' y_i + b' \quad a' = \frac{\Delta_s}{\Delta'}$$
$$\Delta' = N \sum_{i=1}^N y_i^2 - \left(\sum_{i=1}^N y_i \right)^2$$

Si hay una correlación real entre x e y deberá existir una relación entre a, a', b y b' .

$$x_i = \frac{1}{a} y_i - \frac{b}{a} \Rightarrow a' = \frac{1}{a}; b' = -\frac{b}{a} \Rightarrow aa' = 1$$

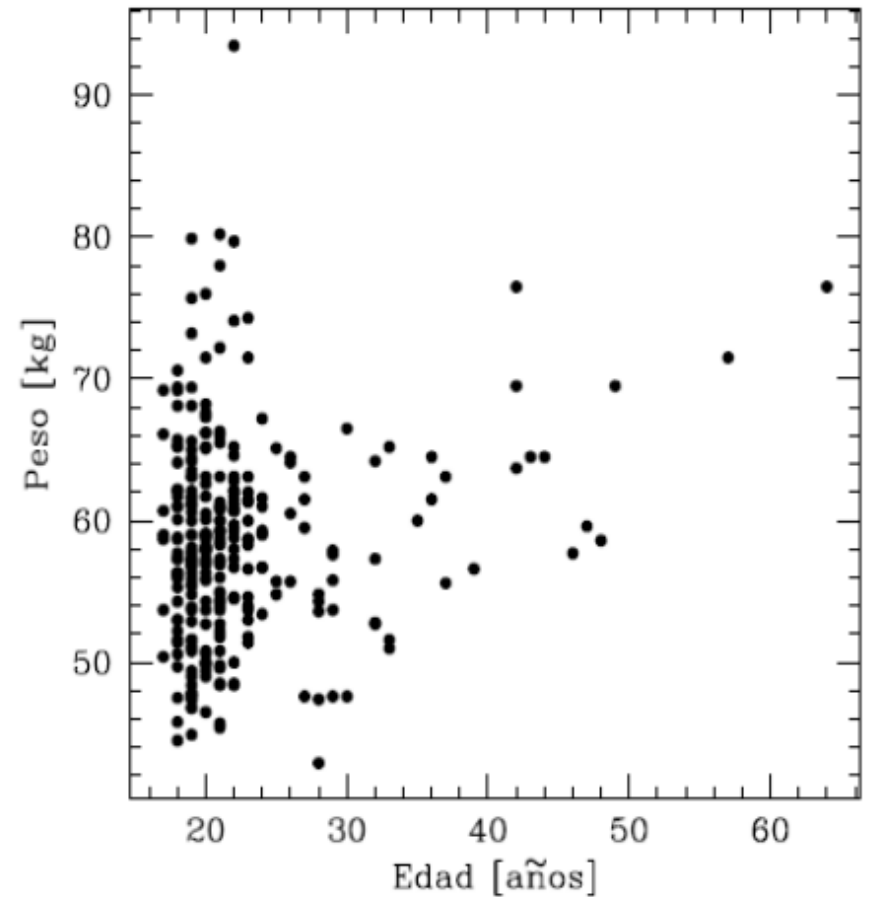
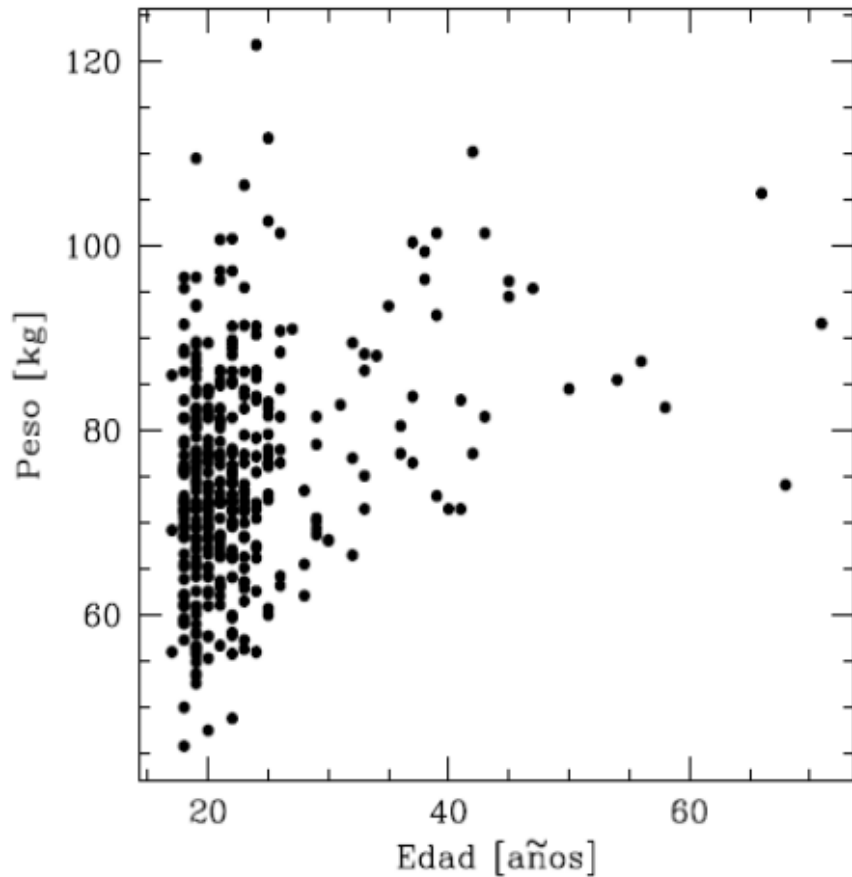
$$aa' = 1$$

Definimos $r = \sqrt{aa'}$ cantidad llamada “coeficiente de correlación lineal”, que nos da una medida experimental del grado de correlación lineal, con valor entre 0 y ± 1 .

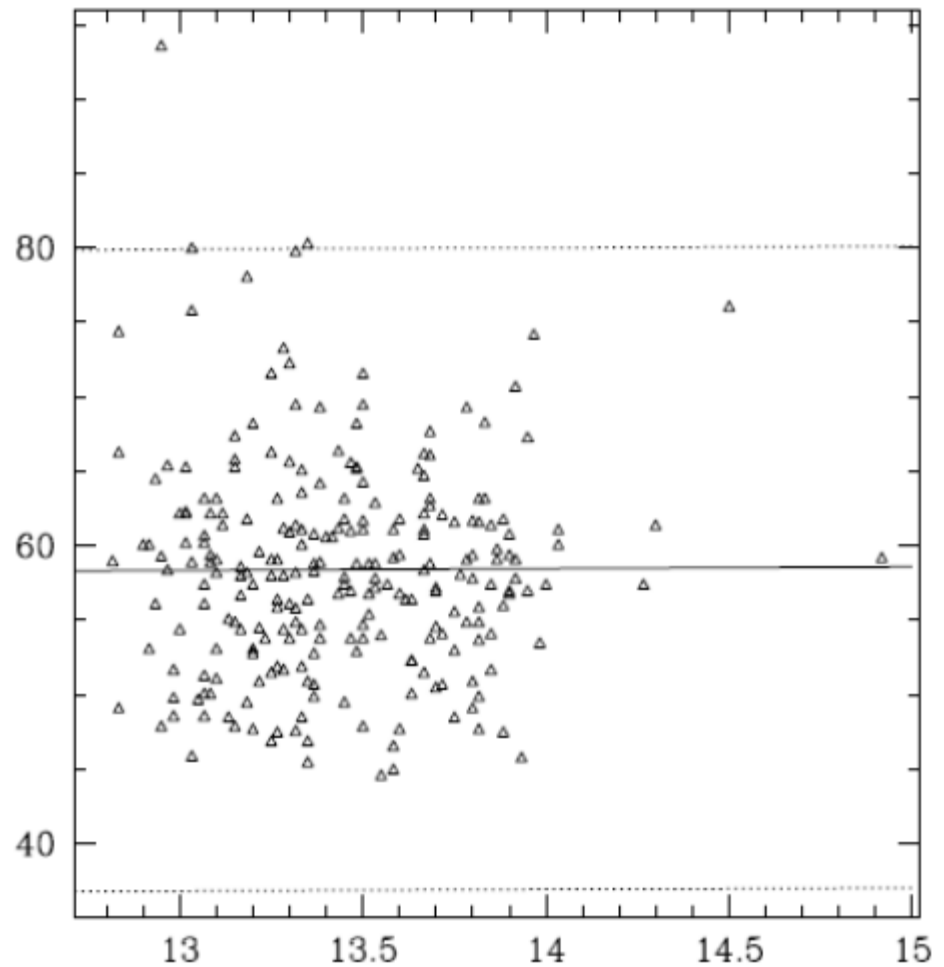
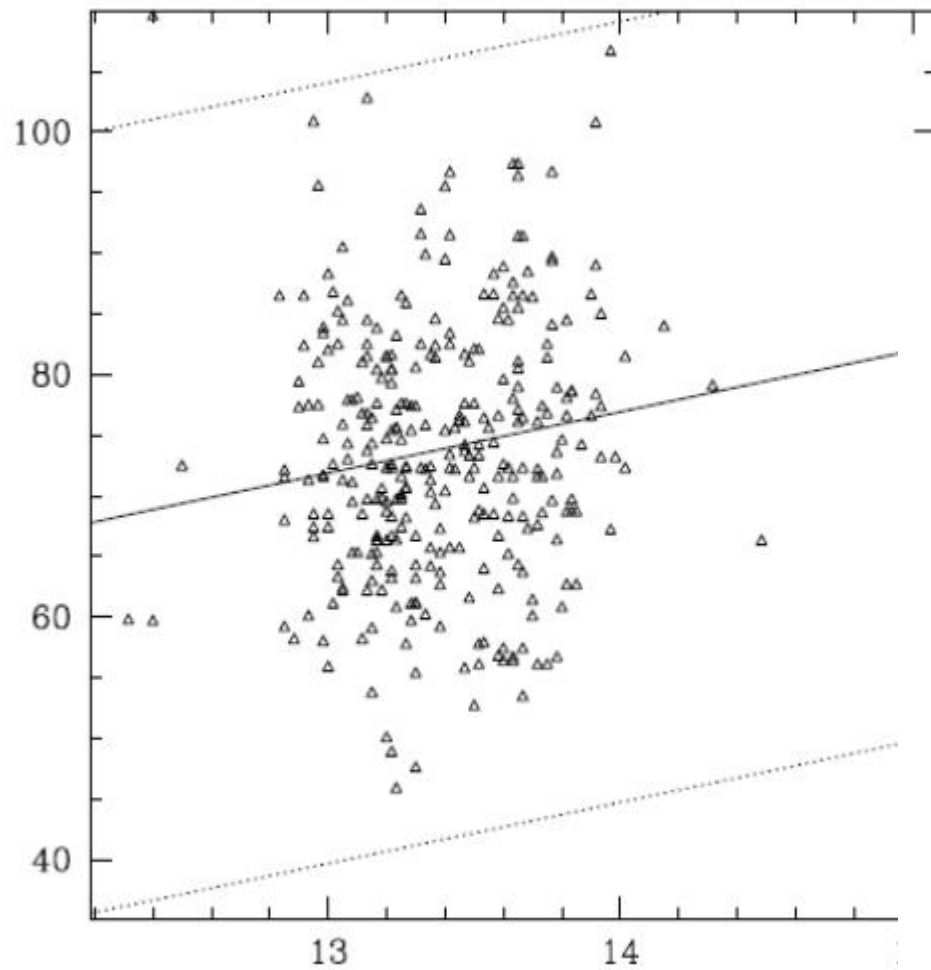
$$r = \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{\sqrt{N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2} \sqrt{N \sum_{i=1}^N y_i^2 - \left(\sum_{i=1}^N y_i \right)^2}}$$

(Raimundo, antes de que tomáramos la raíz cuadrada, el numerador era un cuadrado)

¿Correlaciones casuales?



¿Correlaciones casuales?



Fin de ppt de Clase 9