Estadística Descriptiva.

1. Dem. $-1 \leq r_{xy} \leq 1$ donde $r_{xy} = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$

dem) Consideremos la siguiente función $h(t)$ definida como

$$h(t) = E((X - \mu x)t + (Y - \mu y))^2$$
$$= t^2 \sigma_X^2 + 2t\,Cov(X, Y) + \sigma_Y^2$$

Como $h(t) \geq 0$ y se trata de una función cuadrática

$$(2\,Cov(X, Y))^2 - 4\sigma_X^2 \sigma_Y^2 \leq 0$$

Lo que es equivalente a $-\sigma_X \sigma_Y \leq Cov(X, Y) \leq \sigma_X \sigma_Y$

Que es $-1 \leq r_{xy} \leq 1$

■

2. Dem. que $|r_{xy}| = 1$ si y solo si existen una relacion lineal entre las variables X e Y, i.e., $y_i = \alpha + \beta x_i$ con $\beta \neq 0$.

dem) Sabemos que $|r_{xy}| = 1$ si y solo si el discriminante es cero.
Esto es, si y solo si $h(t)$ tiene una sola raíz.
Pero como $((X - \mu x)t + (Y - \mu y))^2 \geq 0$, $h(t) = 0$ si y solo si

$$\mathbb{P}((X - \mu x)t + (Y - \mu y) = 0) = 1$$

Esto $\mathbb{P}(Y = \alpha + \beta X) = 1$ con $\alpha = \mu_x t + \mu_y$ y $\beta = -t$

donde t es la raíz de $h(t)$.

■

Complejidad Computacional

1. Si f_1 es $O(g_1)$ y f_2 es $O(g_2)$ entonces f_1+f_2 es $O(\max |g_1|, |g_2|)$.

dem) Por definición sabemos que para una función dada $g(n)$

$$O(g(n)) = \{ f(n) : \text{existen ctes. positivas } c \text{ y } n_0 \text{ tq. } 0 \le f(n) \le c g(n) \ \forall n \ge n_0 \}$$

Como f_1 es $O(g_1)$ y f_2 es $O(g_2)$

entonces existen c_1, c_2, n_1 y n_2 ctes. positivas tq

$$0 \le f_1 \le c_1 g_1 \quad \forall n \ge n_1 \quad y \quad 0 \le f_2 \le c_2 g_2 \quad \forall n \ge n_2$$

Sumando ambas expresiones obtenemos:

$$0 \le f_1 + f_2 \le c_1 g_1 + c_2 g_2 \quad \forall n \ge \max\{n_1, n_2\}$$

entonces $c_1 g_1 + c_2 g_2 \ge 0$ por propiedades de valor absoluto

$$-|c_1 g_1 + c_2 g_2| \le c_1 g_1 + c_2 g_2 \le |c_1 g_1 + c_2 g_2|$$

de donde $|c_1 g_1 + c_2 g_2| \le |c_1 g_1| + |c_2 g_2|$

lo que implica $0 \le |f_1| + |f_2| \le |c_1 g_1| + |c_2 g_2|$

esto es, $0 \le |f_1| + |f_2| \le (|c_1| + |c_2|) \max(|g_1|, |g_2|)$ $\quad \forall n \ge \max\{n_1, n_2\}$

como $c = |c_1| + |c_2|$ y $n_0 = \max\{n_1, n_2\}$

se cumple que

$$f_1 + f_2 \text{ es } O(\max(|g_1|, |g_2|))$$

2. Si f_1 es $O(g_1)$ y f_2 es $O(g_2)$ entonces $f_1 f_2$ es $O(g_1 g_2)$.

dem) Usando la definición del ejercicio 1 tenemos que

$$0 \leq f_1 \leq c_1 g_1 \quad \forall n \geq n_1 \quad y \quad 0 \leq f_2 \leq c_2 g_2 \quad \forall n \geq n_2$$

Como ambas desigualdades son positivas, al multiplicar no se afecta la desigualdad

entonces, $0 \leq f_1 f_2 \leq c_1 g_1 c_2 g_2 \quad \forall n \geq max(n_1, n_2)$

luego $0 \leq f_1 f_2 \leq (c_1 c_2) g_1 g_2 \quad \forall n \geq max(n_1, n_2)$

donde $c = c_1 c_2$ y $n_0 = max(n_1, n_2)$

así $f_1 f_2 \leq c g_1 g_2$

por lo tanto $f_1 f_2$ es $O(g_1 g_2)$

3. ¿Cuál es la complejidad del siguiente algoritmo (o)?

```
for (i=0; i<N; i++){
    for (j=0; j<M; j++){
        print(i,j)
    }
}
```

Sol) Sabemos que el tiempo de ejecución de un bucle FOR es el producto del número de iteraciones por la complejidad de las instrucciones del cuerpo del mismo bucle

→ Notemos que el 1er FOR se hara N veces ent $O(N)$
→ Notemos que el 2do FOR se hara M veces ent $O(M)$
→ La instrucción print se hara una sola vez por repetición ent $O(1)$

así $N(O(1)) \cdot M(O(1)) \cdot 1(O(1)) = O(N) \cdot O(M) \cdot O(1) = O(NM)$