*2.- Asuma un modelo de regresión lineal con errores normales y σ conocida. Mostrar que el modelo con AIC mas grande es el modelo C_p Mallows más pequeño.

$\boxed{\text{Dem}}$

Sea $P = \mathcal{P}(\{\beta_0, \beta_1, \ldots, \beta_n\})$ el conjunto potencia de los parámetros de la regresión.

Sea $P^* \in P$, n el tamaño de la muestra y σ^2 la varianza conocida.

Def $\mathcal{L}(P^*) = -n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - P^* X)^2$

$= -n \log \sigma - \frac{RSS(P^*)}{2\sigma^2}$ (Una definición obtenida del libro "Statistical Machine Learning" de Wasserman y Liu) Pág. 213

Def.- $C_p(P^*) = RSS(P^*) + 2|P^*|\sigma^2$ donde $|P^*|$ es la cardinalidad de P^* (Pág 219 del libro "Statistical Machine Learning" de Wasserman y Liu)

Def.- $AIC(P^*) = \mathcal{L}(P^*) - |P^*|$ (Pág 220 del libro "Statistical Machine Learning" de Wasserman y Liu)

Supongamos que $\exists P_0 \in P$ tal que

$$AIC(P_0) > AIC(P^*) \quad \forall P^* \in P$$

$$\Rightarrow -|P_0| - n \log \sigma - \frac{RSS(P_0)}{2\sigma^2} > -n \log \sigma - \frac{RSS(P^*)}{2\sigma^2} - |P^*|$$

$$\Rightarrow -|P_0| - \frac{RSS(P_0)}{2\sigma^2} > - \frac{RSS(P^*)}{2\sigma^2} - |P^*|$$

$$\Rightarrow \frac{RSS(P_0)}{2\sigma^2} + |P_0| < \frac{RSS(P^*)}{2\sigma^2} + |P^*|$$

$$\Rightarrow RSS(P_0) + 2\sigma^2|P_0| < RSS(P^*) + 2\sigma^2|P^*|$$

$$\therefore C_p(P_0) < C_p(P^*) \quad \blacksquare$$